Supporting Information for:

Rapid Assessment of Chemical Complementarity of Ligands for Protein Design

Rokas Petrenas[1,†], Katarzyna Ożga[1,†], Joel J. Chubb[1,†], Andrey V. Romanyuk[1,2], Jennifer J. McManus[3], Graham J. Leggett[4], Nigel S. Scrutton[5,6], Thomas A. A. Oliver[1], and Derek N. Woolfson[1,2,6,*]

[1]School of Chemistry, University of Bristol, Bristol, UK.

[2]Max Planck-Bristol Centre for Minimal Biology, University of Bristol, Bristol, UK.

[3]HH Wills Physics Laboratory, University of Bristol, Bristol, UK.

[4]Department of Chemistry, University of Sheffield, Sheffield, UK.

[5]Manchester Institute of Biotechnology, University of Manchester, UK

[6]Department of Chemistry, University of Manchester, Manchester, UK.

[7]School of Biochemistry, University of Bristol, Medical Sciences Building, Bristol, UK.

[†]These authors contributed equally: Rokas Petrenas, Katarzyna Ożga, Joel J. Chubb.

*To whom correspondence should be addressed: d.n.woolfson@bristol.ac.uk

1 Supplementary Tables

Supplementary Table 1. Computational summary of RASSCCoL. *Since DPH is symmetric, the layers were constrained to 2-fold symmetry. **Active sampling was used to sample 100 000 sequences out of 1 277 192. CPU times are calculated as the number of sequences multiplied by the average packing and docking times measured in the benchmark (Supplementary Figure 1), rounded to the nearest second. For hydrophobic DPH and Nile red, amino acid palette was selected to reflect hydrophobic amino acid preference at **a** or **d** positions within the heptad. For flavin and SN38, amino acid palette was selected to reflect presence/absence of hydrogen bond donor/acceptor within each ligand layer.

Ligand	Scaffold	Layer positions	Layer volumes, Å^3	Amino acid palette	Number of sequences	Single CPU thread time, s
DPH*	8a3k	[9, 77, 51, 118, 80] [44, 111, 16, 84,114]	82 – 192 82 – 192	***a*** = [AVLIG] ***d*** = [AVIG]	3258	1890
Nile red	8a3k	[16,44,111,84,118,51, 114,12,19]	324 – 389	***a*** = [AVLIG] ***d*** = [AVIG]	396 774	273 774
Nile red	8qad	[23,95,167,44,191,116,16,88,160]	259 – 389	***a*** = [AVLIGF]	100 000**	69 000
Lumiflavin	8a3k	[9,77,121] [12,51,118] 16,84,114]	46-86 124-164 98-138	[GAVLISTN] [GAVLISTN] [GAVLI]	1 173 314	746 423
Arcyriaflavin A	8a3k	[16,44,111,84,118,51, 114,12,19]	310 - 387	***a*** = [AVLGTS] ***d*** = [AVG]	332 685	242 860
SN38	8a3k	[12,51,118] [16,84,114] [19,44,87,111]	145-165 145-165 180-200	[GAVLISTN] [GAVLISTN] [GAVLISTN]	51 262	38 447

Supplementary Table 2. AlphaFold2 metrics for all RASSCCoL proteins in this study. *DPH binders were designed with an earlier iteration of the protocol that used all 5 AlphaFold2 models and the default seed ("000").

Name	pLDDT	pTM	bb RMSD to parent, Å
sc-apCC-4-Gly	89.8	0.8	1.1
sc-apCC-4-Ala	88.8	0.8	9.9
sc-apCC-6-Leu	95.2	0.9	0.5
sc-apCC-4-DPH-1	91.2	0.7	0.5
sc-apCC-4-DPH-2	90.5	0.8	10.2
sc-apCC-4-NR-1	92.9	0.9	0.8
sc-apCC-4-NR-2	92.1	0.8	0.9
sc-apCC-4-LMF-1	92.4	0.8	0.4
sc-apCC-4-LMF-2	92.6	0.9	1.1
sc-apCC-4-LMF-3	92.6	0.9	0.8
sc-apCC-4-SN38-1	92.2	0.9	0.9
sc-apCC-4-SN38-2	92.5	0.9	0.9
sc-apCC-4-SN38-3	92.0	0.8	0.9
sc-apCC-4-SN38-4	92.4	0.8	0.9
sc-apCC-4-A7F-1	93.3	0.9	0.6
sc-apCC-4-A7F-2	93.2	0.9	0.7
sc-apCC-6-DPH-NR	95.5	0.9	0.4
sc-apCC-6-NR	95.6	0.9	0.4
sc-apCC-4-DPH-1-sc-apCC-4-NR-1	91.22	0.7	n/a

Supplementary Table 3. Expressed RASSCCoL protein sequences for *in vitro* characterisation in this manuscript.

Name	Scaffold	Sequence
sc-apCC-4-Gly	8a3k	MGSSHHHHHHSSGLVPRGSHM QLEEIAQQGEEGAKQGKKIAWQLKKIAQGEPSAQG QLEEIAQQLEEGAKQGKKGAWQLKKIAQGPDSV QLEEIAQQGEEGAKQGKKIAWQLKKIAQGGTSGG QLEEIAQQLEEGAKQGKKGAWQLKKIAQ
sc-apCC-4-Ala	8a3k	MGSSHHHHHHSSGLVPRGSHM QLEEIAQQLEEAAKQAKKAAWQLKKIAQGEPSAQG QLEEIAQQAEEAAKQAKKIAWQLKKIAQGPDSV QLEEIAQQLEEAAKQAKKAAWQLKKIAQGGTSGG QLEEIAQQAEEAAKQAKKIAWQLKKIAQ
sc-apCC-6-Leu	8qad	MGSSHHHHHHSSGLVPRGSHM LFEEIAQLLEEIAKLLKKIAWLLKKIAQGAQPLEM LFEEIAQLLEEIAKLFKKIAWLLKKIAQTTETKKQGD LFEEIAQLLEEIAKLLKKIAWLLKKIAQGYGDKRT LFEEIAQLLEEIAKLFKKIAWLLKKIAQVAPTQRHRY LFEEIAQLLEEIAKLLKKIAWLLKKIAQGTNSDSDLKS LFEEIAQLLEEIAKLFKKIAWLLKKIAQG
sc-apCC-4-DPH-1	8a3k	MGSSHHHHHHSSGLVPRGSHM QLEEIAQQGEEAAKQAKKIAWQLKKIAQGEPSAQG QLEEIAQQIEEIAKQVKKIAWQLKKIAQGPDSV QLEEIAQQIEEIAKQVKKIAQQLKKIAQGGTSGG QLEEIAQQGEEAAKQAKKIAQQLKKIAQ
sc-apCC-4-DPH-2	8a3k	MGSSHHHHHHSSGLVPRGSHM QLEEIAQQIEEIAKQVKKIAWQLKKIAQGEPSAQG QLEEIAQQIEEIAKQVKKIAWQLKKIAQGPDSV QLEEIAQQGEEAAKQAKKIAQQLKKIAQGGTSGG QLEEIAQQGEEAAKQAKKIAQQLKKIAQ
sc-apCC-4-NR-1	8a3k	MGSSHHHHHHSSGENLYFQSHM QLEEIAQQLEEGAKQAKKAAWQLKKIAQGEPSAQG QLEEIAQQVEEIAKQIKKIAWQLKKIAQGPDSV QLEEIAQQLEEIAKQVKKIAWQLKKIAQGGTSGG QLEEIAQQAEEGAKQAKKIAWQLKKIAQ
sc-apCC-4-NR-2	8a3k	MGSSHHHHHHSSGENLYFQSHM QLEEIAQQLEEGAKQGKKLAWQLKKIAQGEPSAQG QLEEIAQQGEEIAKQGKKIAWQLKKIAQGPDSV QLEEIAQQLEEIAKQVKKIAWQLKKIAQGGTSGG QLEEIAQQIEEGAKQIKKIAWQLKKIAQ
sc-apCC-4-LMF-1	8a3k	MGSSHHHHHHSSGLVPRGSHM QLEEIAQQSEEGAKQGKKIAWQLKKIAQGEPSAQG QLEEIAQQLEEIAKQSKKIAWQLKKIAQGPDSV QLEEIAQQVEEIAKQLKKIAWQLKKIAQGGTSGG QLEEIAQQLEEGAKQAKKTAWQLKKIAQ
sc-apCC-4-LMF-2	8a3k	MGSSHHHHHHSSGENLYFQSHM QLEEIAQQTEEGAKQAKKIAWQLKKIAQGEPSAQG QLEEIAQQLEEIAKQSKKIAWQLKKIAQGPDSV

		QLEEIAQQTEEIAKQVKKIAWQLKKIAQGGTSGG QLEEIAQQLEEGAKQAKKTAWQLKKIAQ
sc-apCC-4-LMF-3	8a3k	MGSSHHHHHHSSGENLYFQSHM QLEEIAQQSEEGAKQAKKIAWQLKKIAQGEPSAQG QLEEIAQQLEEIAKQSKKIAWQLKKIAQGPDSV QLEEIAQQNEEIAKQVKKIAWQLKKIAQGGTSGG QLEEIAQQLEEGAKQSKKTAWQLKKIAQ
sc-apCC-4-A7F-1	8a3k	MGSSHHHHHHSSGLVPRGSHM QLEEIAQQLEEGAKQTKKVAWQLKKIAQGEPSAQG QLEEIAQQVEEIAKQLKKIAWQLKKIAQGPDSV QLEEIAQQLEEIAKQSKKIAWQLKKIAQGGTSGG QLEEIAQQGEEGAKQAKKIAWQLKKIAQ
sc-apCC-4-A7F-2	8a3k	MGSSHHHHHHSSGLVPRGSHM QLEEIAQQLEEGAKQAKKGAWQLKKIAQGEPSAQG QLEEIAQQGEEIAKQLKKIAWQLKKIAQGPDSV QLEEIAQQLEEIAKQVKKIAWQLKKIAQGGTSGG QLEEIAQQTEEGAKQLKKIAWQLKKIAQ
sc-apCC-4-SN38-1	8a3k	MGSSHHHHHHSSGLVPRGSHMTSME QLEEIAEQLEEGAQQAKKNAKQLKKIAKGEPSAQGE QLEEIAEQAEEIAQQSKKIAKQLKKIAKGPDSVE QLEEIAEQLEEIAQQSKKGAKQLKKIAKGGTWSGGE QLEEIAEQAEEGAQQAKKIAKQLKKIAK
sc-apCC-4-SN38-2	8a3k	MGSSHHHHHHSSGLVPRGSHMTSME QLEEIAEQLEEGAQQAKKNAKQLKKIAKGEPSAQGE QLEEIAEQAEEIAQQTKKIAKQLKKIAKGPDSVE QLEEIAEQLEEIAQQSKKGAKQLKKIAKGGTWSGGE QLEEIAEQAEEGAQQAKKIAKQLKKIAK
sc-apCC-4-SN38-3	8a3k	MGSSHHHHHHSSGLVPRGSHMTSME QLEEIAEQLEEGAQQAKKAAKQLKKIAKGEPSAQGE QLEEIAEQGEEIAQQSKKIAKQLKKIAKGPDSVE QLEEIAEQLEEIAQQSKKNAKQLKKIAKGGTWSGGE QLEEIAEQAEEGAQQAKKIAKQLKKIAK
sc-apCC-4-SN38-4	8a3k	MGSSHHHHHHSSGLVPRGSHMTSME QLEEIAEQLEEGAQQAKKNAKQLKKIAKGEPSAQGE QLEEIAEQAEEIAQQSKKIAKQLKKIAKGPDSVE QLEEIAEQLEEIAQQSKKGAKQLKKIAKGGTWSGGE QLEEIAEQAEEGAQQAKKIAKQLKKIAK
sc-apCC-6-DPH-NR	8qad	MGSSHHHHHHSSGLVPRGSHM LFEEIAQLLEEIAKLVKKIAWLVKKIAQGAQPLEM LFEEIAQLAEEIAKLFKKIAWLLKKIAQTTETKKQGD LFEEIAQLLEEIAKLIKKIAWLIKKIAQGYGDKRT LFEEIAQLFEEIAKLFKKIAWLLKKIAQVAPTQRHRY LFEEIAQLLEEIAKLFKKIAWLFKKIAQGTNSDSDLKS LFEEIAQLVEEIAKLFKKIAWLLKKIAQG
sc-apCC-6-NR	8qad	MGSSHHHHHHSSGLVPRGSHM LFEEIAQLLEEIAKLIKKIAWLFKKIAQGAQPLEM LFEEIAQLAEEIAKLFKKIAWLLKKIAQTTETKKQGD LFEEIAQLLEEIAKLFKKIAWLIKKIAQGYGDKRT LFEEIAQLIEEIAKLFKKIAWLLKKIAQVAPTQRHRY LFEEIAQLLEEIAKLFKKIAWLGKKIAQGTNSDSDLKS

		LFEEIAQLVEEIAKLFKKIAWLLKKIAQG
sc-apCC-4-DPH-1-sc-apCC-4-NR-1	2 x 8a3k	MGSSHHHHHHSSGENLYFQSHMM QLEEIAQQGEEAAKQAKKIAWQLKKIAQGEPSAQG QLEEIAQQIEEIAKQVKKIAWQLKKIAQGPDSV QLEEIAQQIEEIAKQVKKIAQQLKKIAQGGTSGG QLEEIAQQGEEAAKQAKKIAQQLKKIAQQLEQIAM QLEEIAQQLEEGAKQGKKLAWQLKKIAQGEPSAQG QLEEIAQQGEEIAKQGKKIAWQLKKIAQGPDSV QLEEIAQQLEEIAKQVKKIAWQLKKIAQGGTSGG QLEEIAQQIEEGAKQIKKIAWQLKKIAQ

Supplementary Table 4. Expressed protein sequences for in-cell characterisation in this manuscript.

Name	Sequence
sc-apCC-4-SN38-1 : mEmerald	MQLEEIAQQLEEGAKQTKKNAWQLKKIAQGEPSAQGQLEEI AQQAEEIAKQSKKIAWQLKKIAQGPDSVQLEEIAQQLEEIA KQAKKGAWQLKKIAQGGTSGGQLEEIAQQAEEGAKQAKKIA WQLKKIAQGSGSGMVSKGEELFTGVVPILVELDGDVNGHKF SVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGV QCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTR AEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHKV YITADKQKNGIKVNFKTRHNIEDGSVQLADHYQQNTPIGDG PVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITLG MDELYK*
NLS : sc-apCC-4-SN38-1 : mEmerald	MPKKKRKVTGSTGSGTMQLEEIAQQLEEGAKQTKKNAWQLK KIAQGEPSAQGQLEEIAQQAEEIAKQSKKIAWQLKKIAQGP DSVQLEEIAQQLEEIAKQAKKGAWQLKKIAQGGTSGGQLEE IAQQAEEGAKQAKKIAWQLKKIAQGSGSGMVSKGEELFTGV VPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGK LPVPWPTLVTTLTYGVQCFARYPDHMKQHDFFKSAMPEGYV QERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDG NILGHKLEYNYNSHKVYITADKQKNGIKVNFKTRHNIEDGS VQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRD HMVLLEFVTAAGITLGMDELYK*
Fyn : sc-apCC-4-SN38-1 : mEmerald	MGCVQCKDKEATKLTEERDGSLNQTSGSGSSRGRSRPPEGT MQLEEIAQQLEEGAKQTKKNAWQLKKIAQGEPSAQGQLEEI AQQAEEIAKQSKKIAWQLKKIAQGPDSVQLEEIAQQLEEIA KQAKKGAWQLKKIAQGGTSGGQLEEIAQQAEEGAKQAKKIA WQLKKIAQGSGSGMVSKGEELFTGVVPILVELDGDVNGHKF SVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGV QCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTR AEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHKV YITADKQKNGIKVNFKTRHNIEDGSVQLADHYQQNTPIGDG PVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITLG MDELYK*
LifeAct : sc-apCC-4-SN38-1 : mEmerald	MGVADLIKKFESISKEETGSTGSGTMQLEEIAQQLEEGAKQ TKKNAWQLKKIAQGEPSAQGQLEEIAQQAEEIAKQSKKIAW QLKKIAQGPDSVQLEEIAQQLEEIAKQAKKGAWQLKKIAQG GTSGGQLEEIAQQAEEGAKQAKKIAWQLKKIAQGSGSGMVS KGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLT LKFICTTGKLPVPWPTLVTTLTYGVQCFARYPDHMKQHDFF KSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIEL KGIDFKEDGNILGHKLEYNYNSHKVYITADKQKNGIKVNFK TRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKL SKDPNEKRDHMVLLEFVTAAGITLGMDELYK*
sc-apCC-4-NR-1 : mEmerald	MQLEEIAQQLEEGAKQAKKAAWQLKKIAQGEPSAQGQLEEI AQQVEEIAKQIKKIAWQLKKIAQGPDSVQLEEIAQQLEEIA KQVKKIAWQLKKIAQGGTSGGQLEEIAQQAEEGAKQAKKIA WQLKKIAQGSGSGMVSKGEELFTGVVPILVELDGDVNGHKF SVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGV QCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTR

	AEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHKV YITADKQKNGIKVNFKTRHNIEDGSVQLADHYQQNTPIGDG PVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITLG MDELYK*
NLS : sc-apCC-4-NR-1 : mEmerald	MPKKKRKVTGSTGSGTMQLEEIAQQLEEGAKQAKKAAWQLK KIAQGEPSAQGQLEEIAQQVEEIAKQIKKIAWQLKKIAQGP DSVQLEEIAQQLEEIAKQVKKIAWQLKKIAQGGTSGGQLEE IAQQAEEGAKQAKKIAWQLKKIAQGSGSGMVSKGEELFTGV VPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGK LPVPWPTLVTTLTYGVQCFARYPDHMKQHDFFKSAMPEGYV QERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDG NILGHKLEYNYNSHKVYITADKQKNGIKVNFKTRHNIEDGS VQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRD HMVLLEFVTAAGITLGMDELYK*
Fyn : sc-apCC-4-NR-1 : mEmerald	MGCVQCKDKEATKLTEERDGSLNQTSGSGSSRGRSRPPEGT MQLEEIAQQLEEGAKQAKKAAWQLKKIAQGEPSAQGQLEEI AQQVEEIAKQIKKIAWQLKKIAQGPDSVQLEEIAQQLEEIA KQVKKIAWQLKKIAQGGTSGGQLEEIAQQAEEGAKQAKKIA WQLKKIAQGSGSGMVSKGEELFTGVVPILVELDGDVNGHKF SVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGV QCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTR AEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHKV YITADKQKNGIKVNFKTRHNIEDGSVQLADHYQQNTPIGDG PVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITLG MDELYK*
LifeAct : sc-apCC-4-NR-1 : mEmerald	MGVADLIKKFESISKEETGSTGSGTMQLEEIAQQLEEGAKQ AKKAAWQLKKIAQGEPSAQGQLEEIAQQVEEIAKQIKKIAW QLKKIAQGPDSVQLEEIAQQLEEIAKQVKKIAWQLKKIAQG GTSGGQLEEIAQQAEEGAKQAKKIAWQLKKIAQGSGSGMVS KGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLT LKFICTTGKLPVPWPTLVTTLTYGVQCFARYPDHMKQHDFF KSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIEL KGIDFKEDGNILGHKLEYNYNSHKVYITADKQKNGIKVNFK TRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKL SKDPNEKRDHMVLLEFVTAAGITLGMDELYK*
sc-apCC-4 : mEmerald	MQLEEIAQQLEEIAKQLKKIAWQLKKIAQGEPSAQGQLEEI AQQLEEIAKQLKKIAWQLKKIAQGPDSVQLEEIAQQLEEIA KQLKKIAWQLKKIAQGGTSGGQLEEIAQQLEEIAKQLKKIA WQLKKIAQGSGSGMVSKGEELFTGVVPILVELDGDVNGHKF SVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGV QCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTR AEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHKV YITADKQKNGIKVNFKTRHNIEDGSVQLADHYQQNTPIGDG PVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITLG MDELYK*
NLS : sc-apCC-4 : mEmerald	MQLEEIAQQLEEIAKQLKKIAWQLKKIAQGEPSAQGQLEEI AQQLEEIAKQLKKIAWQLKKIAQGPDSVQLEEIAQQLEEIA KQLKKIAWQLKKIAQGGTSGGQLEEIAQQLEEIAKQLKKIA WQLKKIAQGSGSGMVSKGEELFTGVVPILVELDGDVNGHKF SVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGV QCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTR

	AEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHKV YITADKQKNGIKVNFKTRHNIEDGSVQLADHYQQNTPIGDG PVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITLG MDELYK*
Fyn : sc-apCC-4 : mEmerald	MGCVQCKDKEATKLTEERDGSLNQTSGSGSSRGRSRPPEGT MQLEEIAQQLEEIAKQLKKIAWQLKKIAQGEPSAQGQLEEI AQQLEEIAKQLKKIAWQLKKIAQGPDSVQLEEIAQQLEEIA KQLKKIAWQLKKIAQGGTSGGQLEEIAQQLEEIAKQLKKIA WQLKKIAQGSGSGMVSKGEELFTGVVPILVELDGDVNGHKF SVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGV QCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTR AEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHKV YITADKQKNGIKVNFKTRHNIEDGSVQLADHYQQNTPIGDG PVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITLG MDELYK*
LifeAct : sc-apCC-4 : mEmerald	MGVADLIKKFESISKEETGSTGSGTMQLEEIAQQLEEIAKQ LKKIAWQLKKIAQGEPSAQGQLEEIAQQLEEIAKQLKKIAW QLKKIAQGPDSVQLEEIAQQLEEIAKQLKKIAWQLKKIAQG GTSGGQLEEIAQQLEEIAKQLKKIAWQLKKIAQGSGSGMVS KGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLT LKFICTTGKLPVPWPTLVTTLTYGVQCFARYPDHMKQHDFF KSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIEL KGIDFKEDGNILGHKLEYNYNSHKVYITADKQKNGIKVNFK TRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKL SKDPNEKRDHMVLLEFVTAAGITLGMDELYK*
sc-apCC-4-SN38-1	MQLEEIAQQLEEGAKQTKKNAWQLKKIAQGEPSAQGQLEEI AQQAEEIAKQSKKIAWQLKKIAQGPDSVQLEEIAQQLEEIA KQAKKGAWQLKKIAQGGTSGGQLEEIAQQAEEGAKQAKKIA WQLKKIAQGS
NLS : sc-apCC-4-SN38-1	MPKKKRKVTGSTGSGTMQLEEIAQQLEEGAKQTKKNAWQLK KIAQGEPSAQGQLEEIAQQAEEIAKQSKKIAWQLKKIAQGP DSVQLEEIAQQLEEIAKQAKKGAWQLKKIAQGGTSGGQLEE IAQQAEEGAKQAKKIAWQLKKIAQGS
Fyn : sc-apCC-4-SN38-1	MGCVQCKDKEATKLTEERDGSLNQTSGSGSSRGRSRPPEGT MQLEEIAQQLEEGAKQTKKNAWQLKKIAQGEPSAQGQLEEI AQQAEEIAKQSKKIAWQLKKIAQGPDSVQLEEIAQQLEEIA KQAKKGAWQLKKIAQGGTSGGQLEEIAQQAEEGAKQAKKIA WQLKKIAQGS
LifeAct : sc-apCC-4-SN38-1	MGVADLIKKFESISKEETGSTGSGTMQLEEIAQQLEEGAKQ TKKNAWQLKKIAQGEPSAQGQLEEIAQQAEEIAKQSKKIAW QLKKIAQGPDSVQLEEIAQQLEEIAKQAKKGAWQLKKIAQG GTSGGQLEEIAQQAEEGAKQAKKIAWQLKKIAQGS

Supplementary Table 5. Experimentally determined dissociation constants K_D (μM) and derived from AutoDock Vina scores (in parentheses) for all proteins in this study.

Name	DPH	Nile red	Lumiflavin	Arcyriaflavin A	SN38
sc-apCC-4-Ala	> 50 (600)	> 20 (-)	no binding (-)	1.8 ± 0.3 (-)	no binding (-)
sc-apCC-6-Leu	6.1 ± 1.9 (0.03)	> 20 (2)			
sc-apCC-4-DPH-1	1.3 ± 0.1 (0.02)	> 30 (20)	no binding (80)	no binding (2)	no binding (-)
sc-apCC-4-DPH-2	2.0 ± 0.2 (0.01)				
sc-apCC-4-NR-1	> 15 (10)	2.4 ± 0.1 (0.04)	no binding (0.05)	0.050 ± 0.020 (0.3)	no binding (30)
sc-apCC-4-NR-2		26 ± 3 (0.007)			
sc-apCC-4-LMF-1	> 70 (10)	1.2 ± 0.1 (0.5)	> 50 (0.01)	0.13 ± 0.02 (0.004)	no binding (32)
sc-apCC-4-LMF-2			no binding (0.008)		
sc-apCC-4-LMF-3			no binding (0.002)		
sc-apCC-4-A7F-1	no binding (10)	2.0 ± 0.2 (0.2)	9.6 ± 0.2 (0.05)	0.11 ± 0.02 (0.001)	no binding (2)
sc-apCC-4-A7F-2				6.5 ± 0.3 (0.002)	
sc-apCC-4-SN38-1	4.4 ± 0.7 (0.3)	0.22 ± 0.04 (2)	> 90 (0.4)	> 20 (0.02)	7.8 ± 0.6 (100)
sc-apCC-4-SN38-2					no binding (50)
sc-apCC-4-SN38-3					no binding (1)
sc-apCC-4-SN38-4					no binding (100)
sc-apCC-6-DPH-NR	1.9 ± 0.2 (0.06)	2.2 ± 0.2 (0.01)			
sc-apCC-6-NR	> 20 (5)	0.11 ± 0.04 (0.04)			
sc-apCC-4-DPH-1-sc-apCC-4-NR-1	1.3 ± 0.4 (0.02)	0.53 ± 0.05 (0.01)			

Supplementary Table 6. Crystallisation conditions for crystal structures in this study.

Protein	Condition name	Condition
sc-apCC-4-DPH-2	JCSG-plus™ D9	0.17 M Ammonium sulfate, 25.5% w/v Polyethylene glycol 4,000, 15% v/v Glycerol
sc-apCC-4-NR-1	Pact Premier™ D2	0.1 M MMT buffer pH 5.0, 25 % w/v PEG 1500
sc-apCC-4-LMF-1	JCSG-plus™ A12	0.2 M Potassium nitrate, 20% w/v Polyethylene glycol 3,350
sc-apCC-4-A7F-1	Pact Premier™ C7	0.2 M sodium chloride, 0.1 M HEPES pH 7.0, 20 % w/v PEG 6000
sc-apCC-4-SN38-1	Pact Premier™ A3	0.1 M SPG buffer pH 6.0, 25 % w/v PEG 1500
sc-apCC-6-DPH-NR	Pact Premier™ B11	0.01 M zinc chloride, 0.1 M MES pH 6.0, 20 % w/v PEG 6000

Supplementary Table 7. Merging and refinement statistics for crystal structures in this study.

	sc-apCC-4-DPH-2	sc-apCC-4-NR-1	sc-apCC-4-LMF-1
PDB id	9R1N	9R1J	9R1M
Wavelength	0.6199 Å	0.6199 Å	0.9999 Å
Resolution range	42.2 - 2.4 (2.486 - 2.4)	27.82 - 1.702 (1.763 - 1.702)	31.33 - 2.4 (2.486 - 2.4)
Space group	P 65 2 2	P 1 21 1	P 1 21 1
Unit cell	68.1003 68.1003 120.793 90 90 120	27.662 38.1608 56.6897 90 101.082 90	28.047 38.26 55.455 90 100.172 90
Total reflections	636957 (64114)	73226 (3364)	8829 (822)
Unique reflections	6975 (674)	12141 (904)	4572 (431)
Multiplicity	91.3 (95.1)	6.0 (3.7)	1.9 (1.9)
Completeness (%)	99.89 (100.00)	94.02 (70.06)	98.62 (97.07)
Mean I/sigma(I)	18.12 (1.89)	25.48 (5.36)	3.90 (1.45)
Wilson B-factor	52.29	21.54	50.63
R-merge	0.2364 (2.615)	0.05404 (0.1202)	0.07497 (0.4823)
R-meas	0.2377 (2.629)	0.05866 (0.1388)	0.106 (0.682)
R-pim	0.02488 (0.2674)	0.0226 (0.06724)	0.07497 (0.4823)
CC1/2	1 (0.895)	0.998 (0.983)	0.984 (0.612)
CC*	1 (0.972)	1 (0.996)	0.996 (0.871)
Reflections used in refinement	6969 (674)	12129 (901)	4570 (430)
Reflections used for R-free	345 (32)	610 (49)	235 (19)
R-work	0.2267 (0.2737)	0.1905 (0.2137)	0.2286 (0.3102)
R-free	0.2422 (0.2908)	0.1881 (0.2109)	0.2353 (0.2515)
CC(work)	0.951 (0.851)	0.961 (0.911)	0.956 (0.728)
CC(free)	0.943 (0.801)	0.944 (0.948)	0.912 (0.856)
Number of non-hydrogen atoms	973	1104	821
macromolecules	943	973	821
ligands	23	14	0

solvent	7	125	0
Protein residues	131	132	129
Nucleic acid bases			
RMS(bonds)	0.007	0.005	0.004
RMS(angles)	0.79	0.52	0.55
Ramachandran favored (%)	100	99.23	99.21
Ramachandran allowed (%)	0	0.77	0.79
Ramachandran outliers (%)	0	0	0
Rotamer outliers (%)	2.6	0	0
Clashscore	5.17	3.59	5.77
Average B-factor	61.18	27.89	63.97
macromolecules	61.03	26.72	63.97
ligands	69.25	43.69	-
solvent	54.83	36.28	-
Number of TLS groups	4	1	5

52

	sc-apCC-4-A7F-1	sc-apCC-4-SN38-1	sc-apCC-6-DPH-NR
PDB id	9R1K	9R1L	9R1O
Wavelength	0.6199 Å	0.6199 Å	0.9762 Å
Resolution range	27.91 - 2.2 (2.279 - 2.2)	27.24 - 1.85 (1.916 - 1.85)	45.66 - 3.229 (3.344 - 3.229)
Space group	P 1 21 1	P 1 21 1	P 1 21 1
Unit cell	27.5551 37.8603 57.0232 90 101.821 90	27.7628 37.5806 54.86 90 101.16 90	58.699 59.583 64.307 90 96.293 90
Total reflections	40693 (4149)	42875 (4053)	12924 (1169)
Unique reflections	5977 (598)	9615 (912)	7018 (654)
Multiplicity	6.8 (6.9)	4.5 (4.4)	1.8 (1.8)

Completeness (%)	99.85 (99.83)	99.73 (98.59)	96.97 (93.97)
Mean I/sigma(I)	9.41 (2.25)	14.39 (1.30)	3.47 (1.43)
Wilson B-factor	41.85	35.34	58.96
R-merge	0.08956 (0.4694)	0.05523 (0.4269)	0.1003 (0.4693)
R-meas	0.097 (0.5074)	0.06274 (0.4868)	0.1419 (0.6637)
R-pim	0.03696 (0.1915)	0.02923 (0.2306)	0.1003 (0.4693)
CC1/2	0.998 (0.932)	0.999 (0.876)	0.995 (0.624)
CC*	0.999 (0.982)	1 (0.966)	0.999 (0.877)
Reflections used in refinement	5970 (597)	9600 (907)	7004 (654)
Reflections used for R-free	298 (26)	434 (40)	322 (35)
R-work	0.2175 (0.2449)	0.2001 (0.2885)	0.2652 (0.3194)
R-free	0.2530 (0.4221)	0.2142 (0.2839)	0.2853 (0.3528)
CC(work)	0.963 (0.906)	0.970 (0.854)	0.924 (0.739)
CC(free)	0.901 (0.872)	0.942 (0.892)	0.772 (0.570)
Number of non-hydrogen atoms	898	1020	2860
macromolecules	881	989	2847
ligands	4	17	13
solvent	13	24	0
Protein residues	126	134	389
Nucleic acid bases			
RMS(bonds)	0.005	0.003	0.001
RMS(angles)	0.62	0.57	0.26
Ramachandran favored (%)	100	99.22	99.45
Ramachandran allowed (%)	0	0.78	0.55
Ramachandran outliers (%)	0	0	0
Rotamer outliers (%)	3.23	0	0
Clashscore	4.7	3.58	3.51

Average B-factor	53.81	46.45	47.04
macromolecules	53.82	46.28	47.07
ligands	61.14	59.44	39.56
solvent	50.95	49.73	-
Number of TLS groups	1	3	2

53

54

Supplementary Table 8. Spectral properties of Nile red and Nile blue bound to sc-apCC-4-SN38-1. Absolute quantum yield was measured with an integrating sphere (Nile red λ_{ex} = 510 nm, Nile blue λ_{ex} = 530 nm). Extinction coefficient was derived through comparison with a known standard – Nile red and Nile blue in methanol.

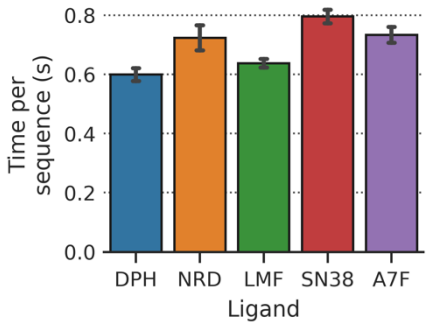
	Φ	τ	ε	λ_{abs} max	λ_{em} max
Nile red	0.90 ± 0.02	4.8 ± 0.2 ns	52 400 $cm^{-1} M^{-1}$ at 552 nm	559 nm	607 nm
Nile blue	0.74 ± 0.02	4.2 ± 0.2 ns	34 400 $cm^{-1} M^{-1}$ at 627 nm	620 nm	645 nm

Supplementary Table 9. Simultaneous FRET lifetime fits from DPH and Nile red channels after excitation at 352 nm. Floated parameters in bold. Fluorescence decays were fitted to bi-exponential decay equations while keeping average DPH and Nile red lifetimes fixed. DPH population A_2 is not involved in FRET; the ratio $A_1/(A_1+A_2)$ represents the fraction of DPH-bound proteins undergoing FRET.

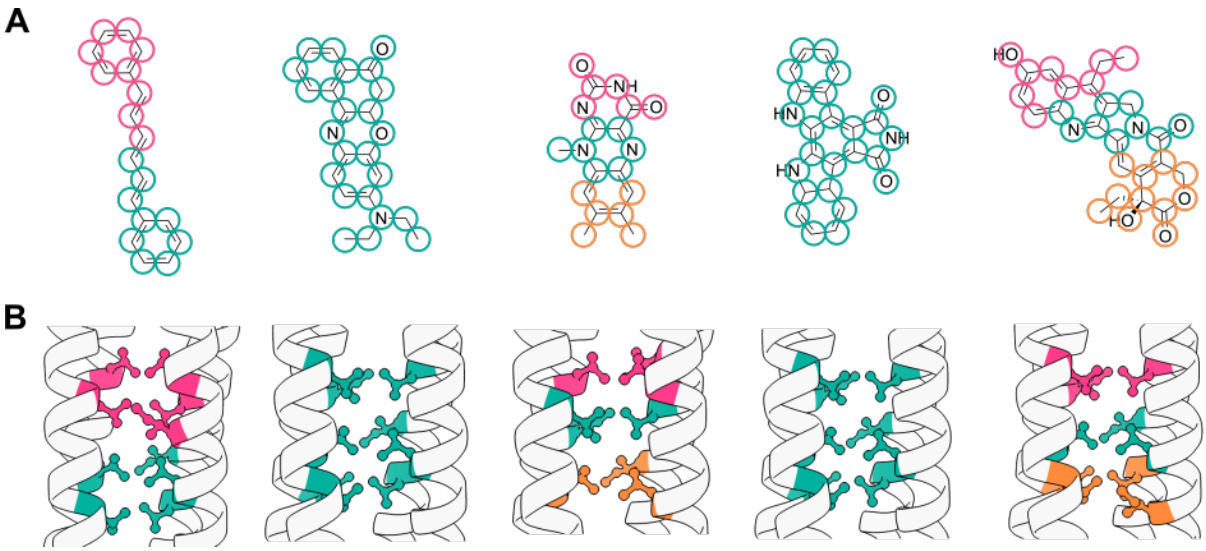
A_1(DPH)	**0.24**
A_2(DPH)	**0.80**
A_2(NR)	**3.2**
t_{FRET}	**2.0 ± 0.2 ns**
t_{DPH}	7.6 ± 0.2 ns
t_{NR}	5.2 ± 0.2 ns

2 Supplementary Figures

2.1 Computational design of 4-helix binders

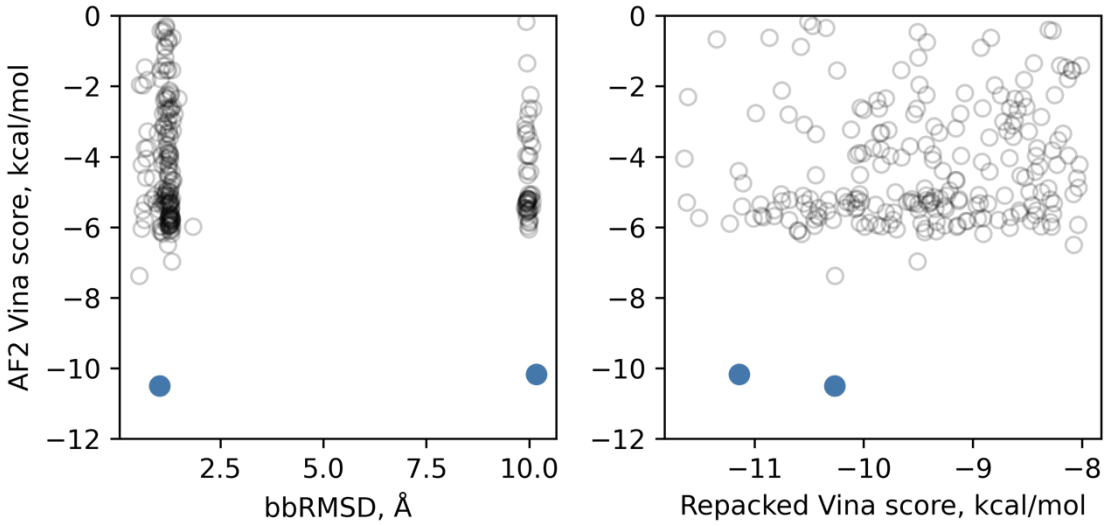


Supplementary Figure 1. CPU benchmarking for backbone sequence packing and ligand docking. The time required to pack a single sequence onto sc-apCC-4 using FASPR and dock a ligand using AutoDock Vina with low sampling parameters was measured on a single thread of a Xeon Gold 6226R processor. A total of 100 sequences were tested per ligand. Bar plots represent the mean time for 1 packing and docking cycle, with error bars denoting one standard deviation. The observed times (in seconds per CPU) were as follows: DPH: 0.58 (± 0.02), Nile red (NRD): 0.69 (± 0.04), lumiflavin (LMF): 0.64 (± 0.01), SN38: 0.75 (± 0.02) and arcyriaflavin A (A7F): 0.73 (± 0.03).

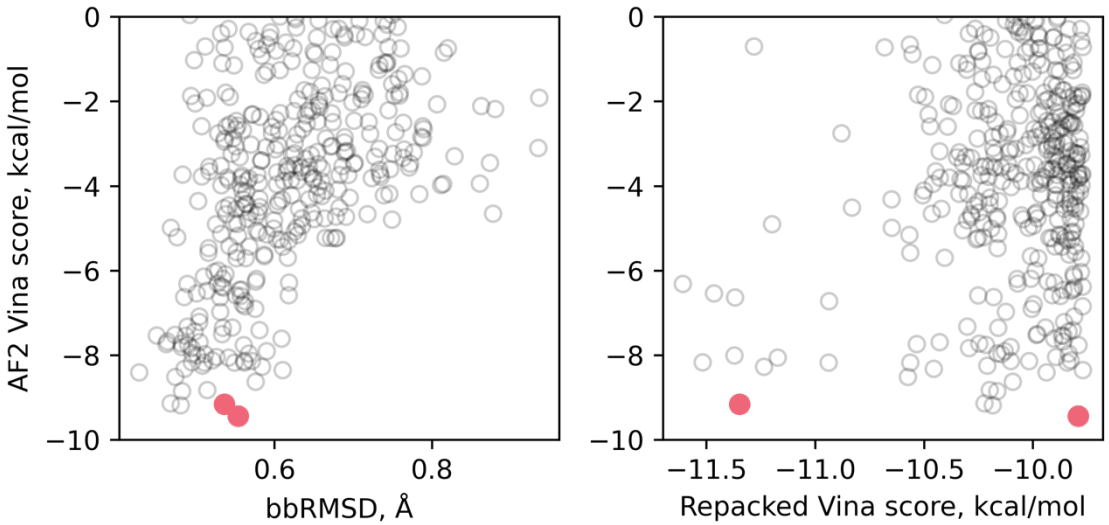


Supplementary Figure 2. Binding site design approach (A) Ligands used in the RASSCCoL pipeline—DPH, Nile red, lumiflavin, arcyriaflavin A, and SN38 (left to right)—divided into colored sections that align with layers in the protein scaffold. **(B)** 4-Helix bundle protein scaffolds are shown

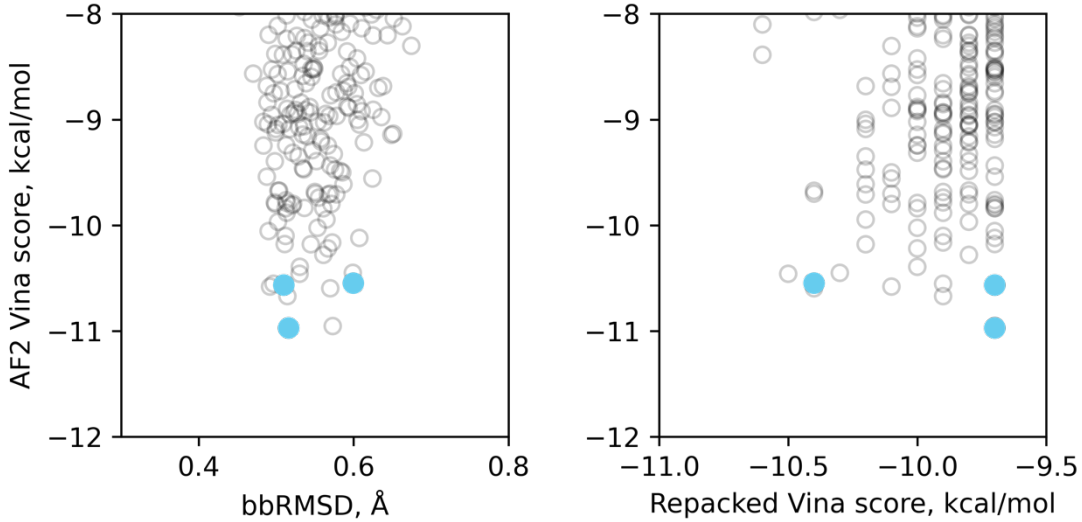
directly below the corresponding ligands, with binding pocket residues candidates represented as ball-and-stick and colored to match the ligand layers. The length of each binding pocket matches the length of its corresponding ligand, allowing for a layer-based design strategy.



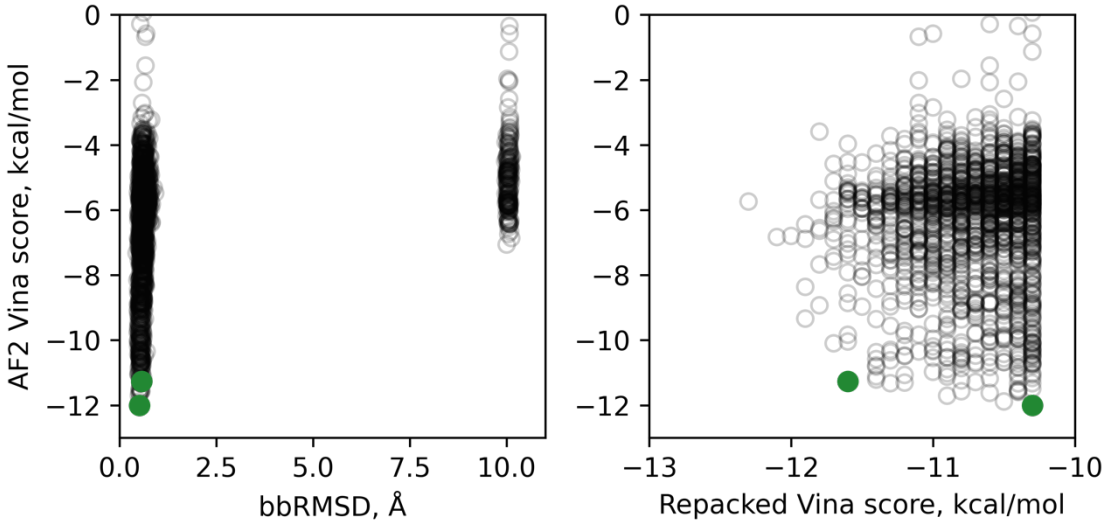
Supplementary Figure 3. Sequence selection for DPH binders based on AutoDock Vina scores. Blue circles denote experimentally characterised designs. Top 270 sequences after repacking (< -8.0 kcal/mol) were analysed with AlphaFold2.



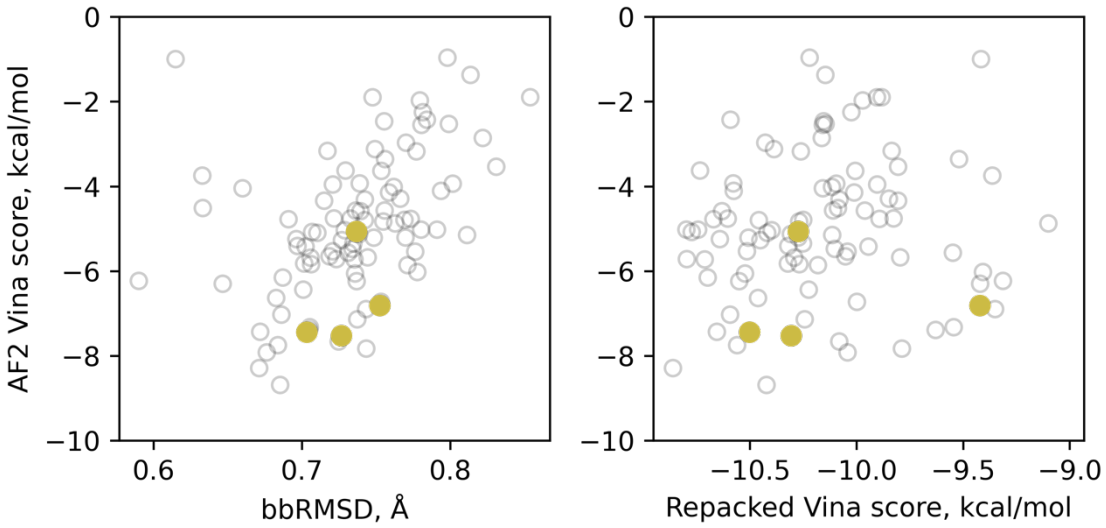
Supplementary Figure 4. Sequence selection for Nile red binders based on AutoDock Vina scores. Red circles denote experimentally characterised designs. Top 600 sequences after repacking (< -9.8 kcal/mol) were analysed with AlphaFold2.



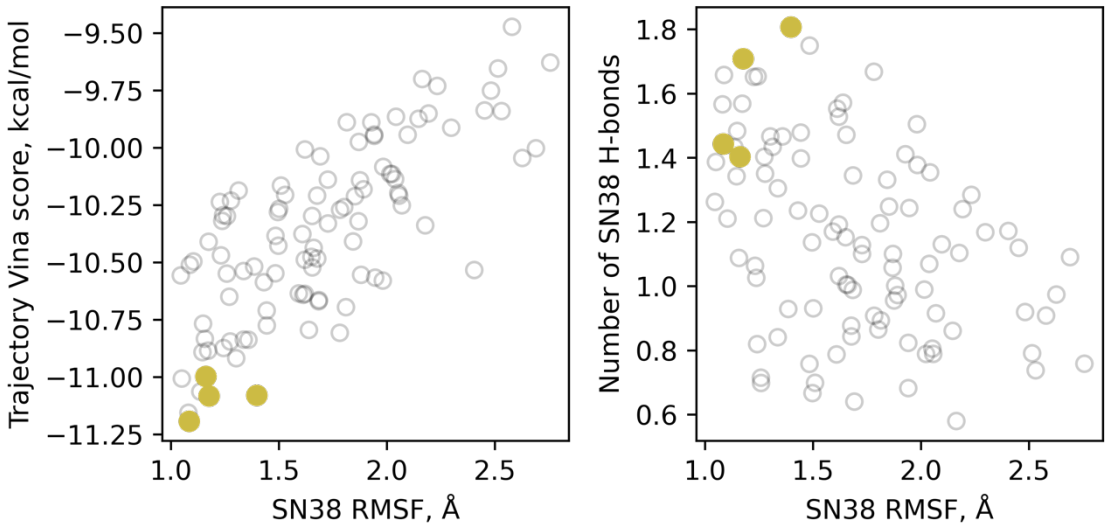
Supplementary Figure 5. Sequence selection for flavin binders based on AutoDock Vina scores. Cyan circles denote experimentally characterised designs. Top 359 sequences after repacking (< -9.6 kcal/mol) were analysed with AlphaFold2.



Supplementary Figure 6. Sequence selection for arcyriaflavin A binders based on AutoDock Vina scores. Green circles denote experimentally characterised designs. Top 1900 sequences after repacking (< -10.2 kcal/mol) were analysed with AlphaFold2.

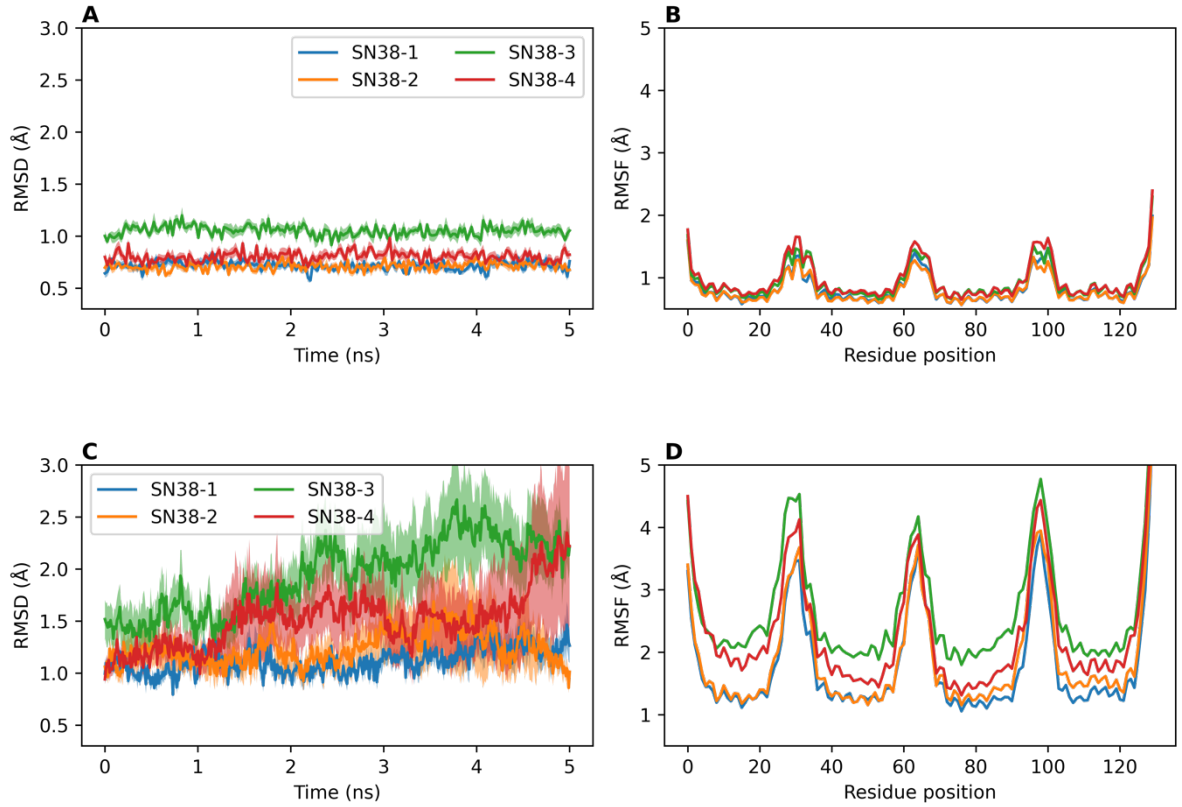


Supplementary Figure 7. Sequence selection for SN38 binders based on AutoDock Vina scores. Yellow circles denote experimentally characterised designs. No design passed the filtering threshold after being folded with AlphaFold2 based on DPH, Nile red, flavin or arcyriaflavin A designs (AF2+Viina score < -9.0 kcal/mol, bbRMSD < 0.6 Å). To better account for ligand-protein hydrogen bonding and rotamer flexibility, we used an additional filtering step with ultrafast MD simulations, see Figures S6&S7.

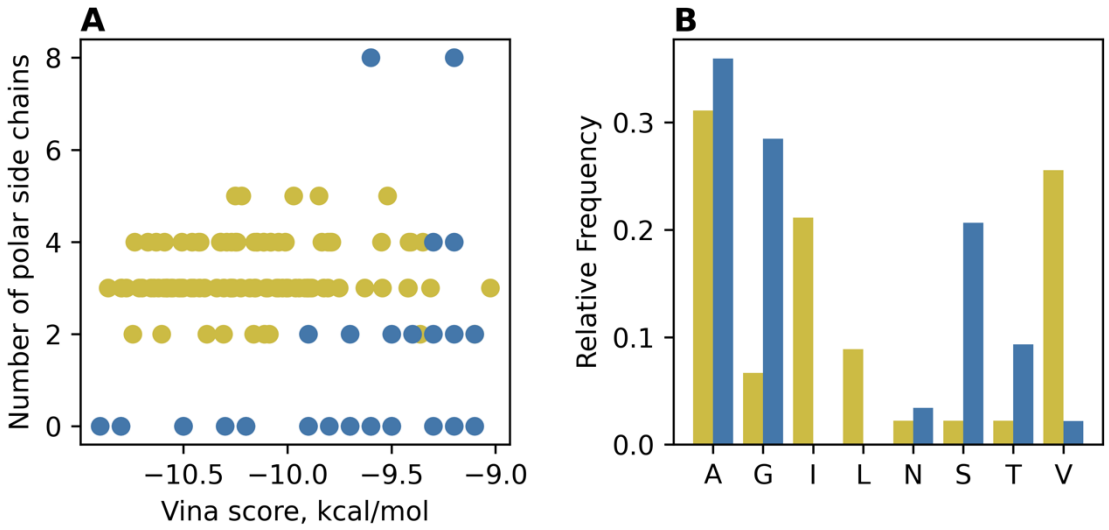


Supplementary Figure 8. Sequence selection for SN38 binders based on ultrafast 10 x 5 ns MD simulations at 400K. RMSF: root-mean-square-fluctuation of SN38. The system with implicit water was minimised, heated to 400 K over 400 ps, and equilibrated for 1.6 ns prior to a 5 ns production run. To enhance sampling, 10 independent replicates of the minimization, equilibration and production runs

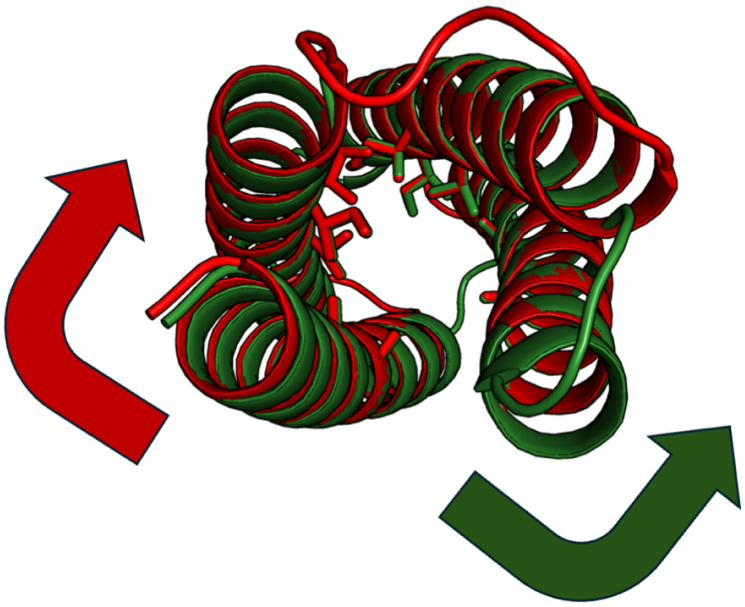
were performed. Weak positional restraints of 20 cal $mol^{-1}Å^{-2}$ were applied to the backbone atoms to maintain simulation stability at elevated temperatures [1].



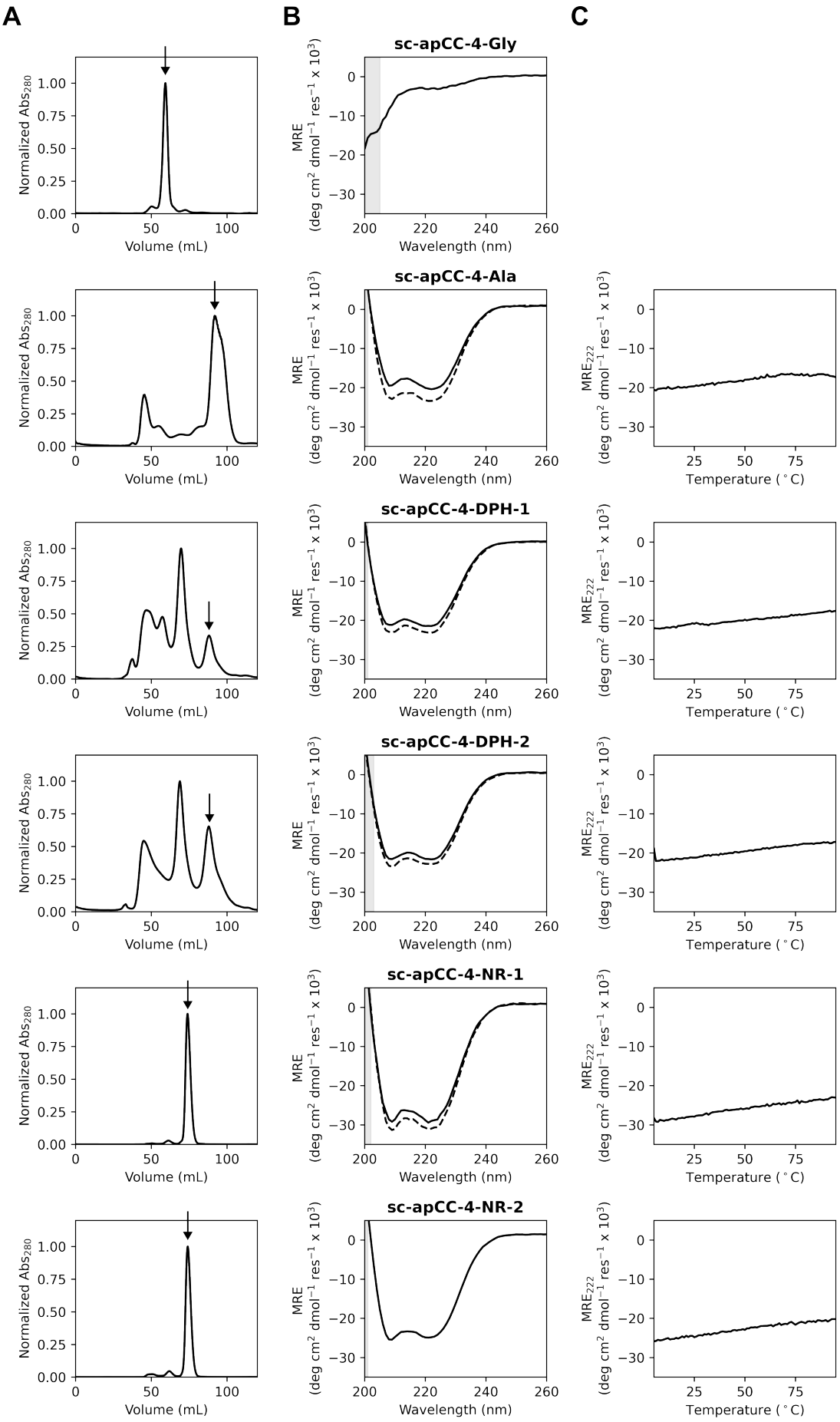
Supplementary Figure 9. Selected SN38 binder binding-site backbone RMSD and Cα RMSF throughout simulations with (A&B) and without (C&D) the 20 cal $mol^{-1}Å^{-2}$ restraint. Weak backbone restraints allow accelerated sampling of the binding site and ligand motion while preventing the structure from undergoing large conformational changes at elevated temperatures [1].



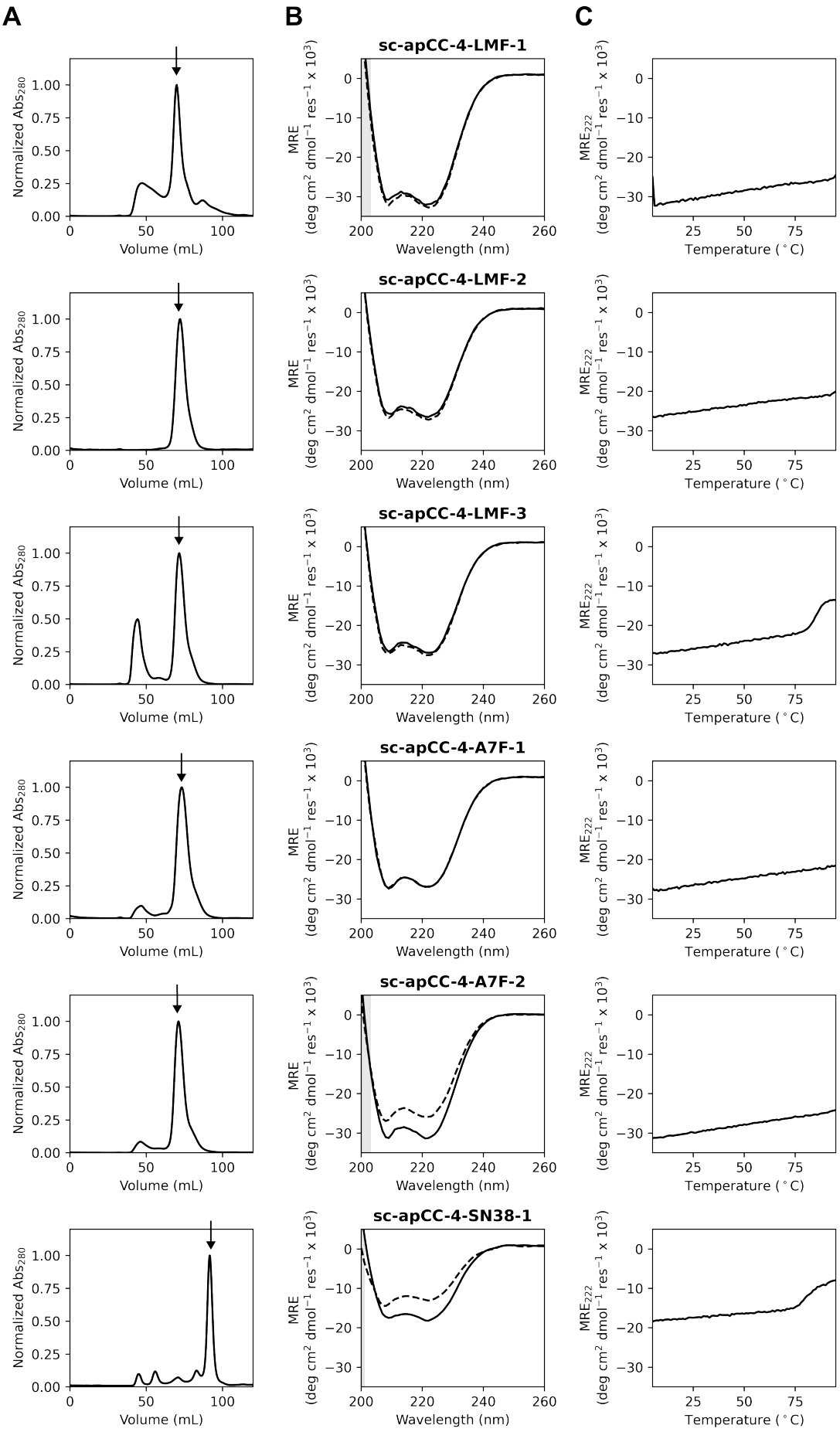
Supplementary Figure 10. Enrichment of polar side chains for SN38 (the most polar ligand, yellow) compared to the DPH (most hydrophobic ligand, blue). **(A)** AutoDock Vina scores for sequences with varying number of polar side chains. For DPH, the best scoring sequences had no polar side chains. For SN38, sequences with binding sites with 2 – 4 polar side chains returned the best scores. **(B)** Frequency of binding-site residues in sequences with <-9.5 kcal/mol Vina score. In both cases, 10 positions were designed in the 4-helix bundle, and the amino acid pallet was limited to A,G,I,L,N,S,T,V.



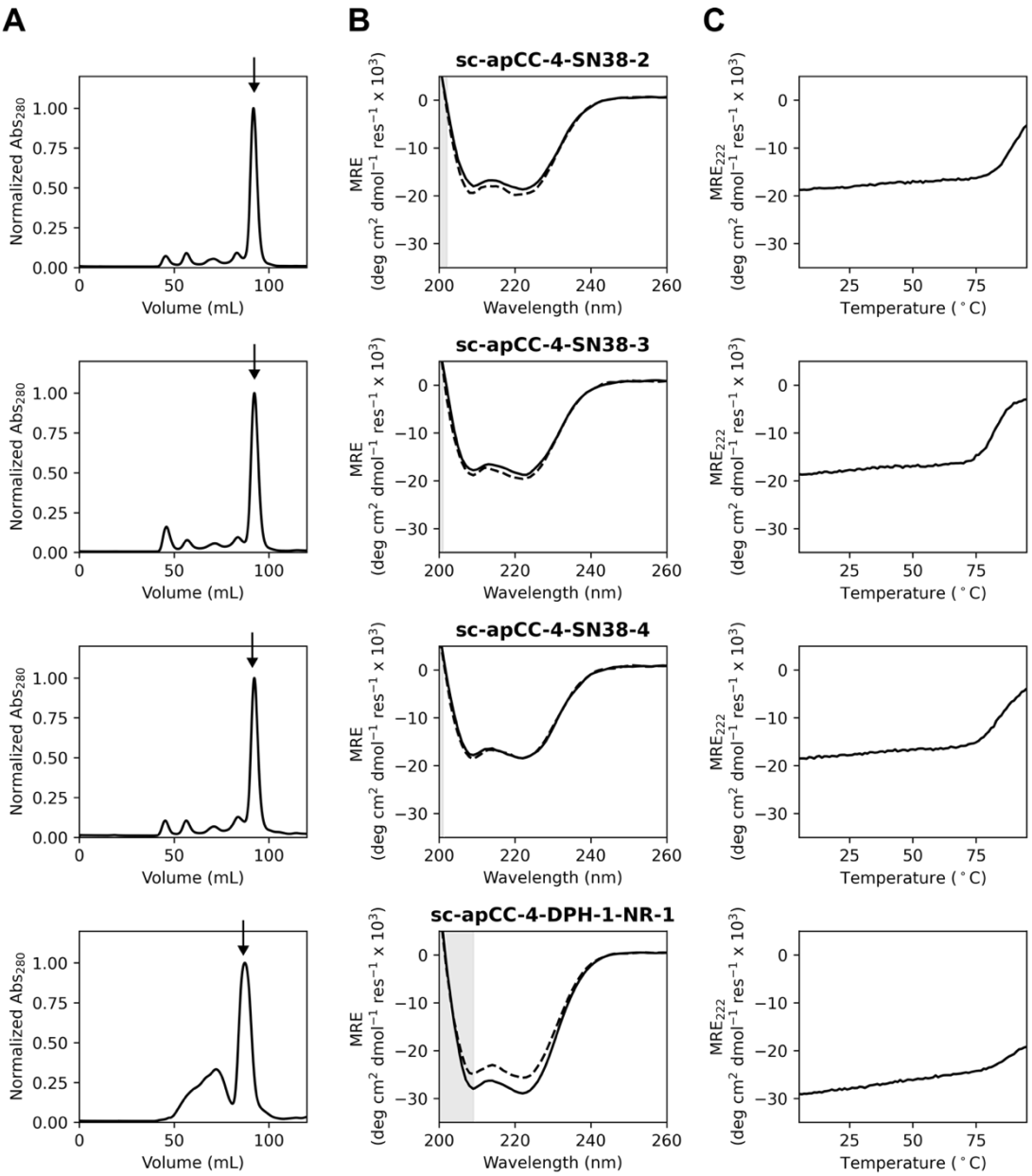
Supplementary Figure 11. AlphaFold2 predicted topology swap by binding site rotation, sc-apCC-4-DPH-2 (clockwise, red) and sc-apCC-4-DPH-1 (anticlockwise, green, same as parent 8a3k). The direction of the topology is shown by arrows and binding site side chains shown as sticks. Despite the global structural change (backbone RMSD = 10.3 Å), the binding site is retained (heavy atom RMSD = 0.2 Å). The topology swap restores optimal knobs-into-holes packing by placing all **Ile** at **d** within the same side of the protein.


A
B
C
sc-apCC-4-Gly
sc-apCC-4-Ala
sc-apCC-4-DPH-1
sc-apCC-4-DPH-2
sc-apCC-4-NR-1
sc-apCC-4-NR-2
Normalized Abs280
Volume (mL)
MRE
(deg cm2 dmol−1 res−1 x 10^3)
Wavelength (nm)
MRE222
Temperature (°C)

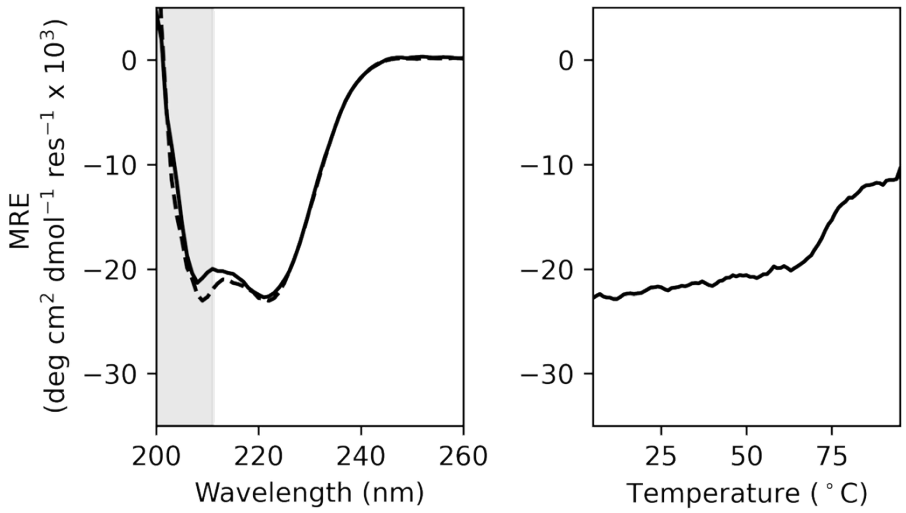
143


A
B
C
sc-apCC-4-LMF-1
sc-apCC-4-LMF-2
sc-apCC-4-LMF-3
sc-apCC-4-A7F-1
sc-apCC-4-A7F-2
sc-apCC-4-SN38-1
Normalized Abs280
Volume (mL)
MRE
(deg cm2 dmol−1 res−1 x 10³)
Wavelength (nm)
MRE222
Temperature (°C)

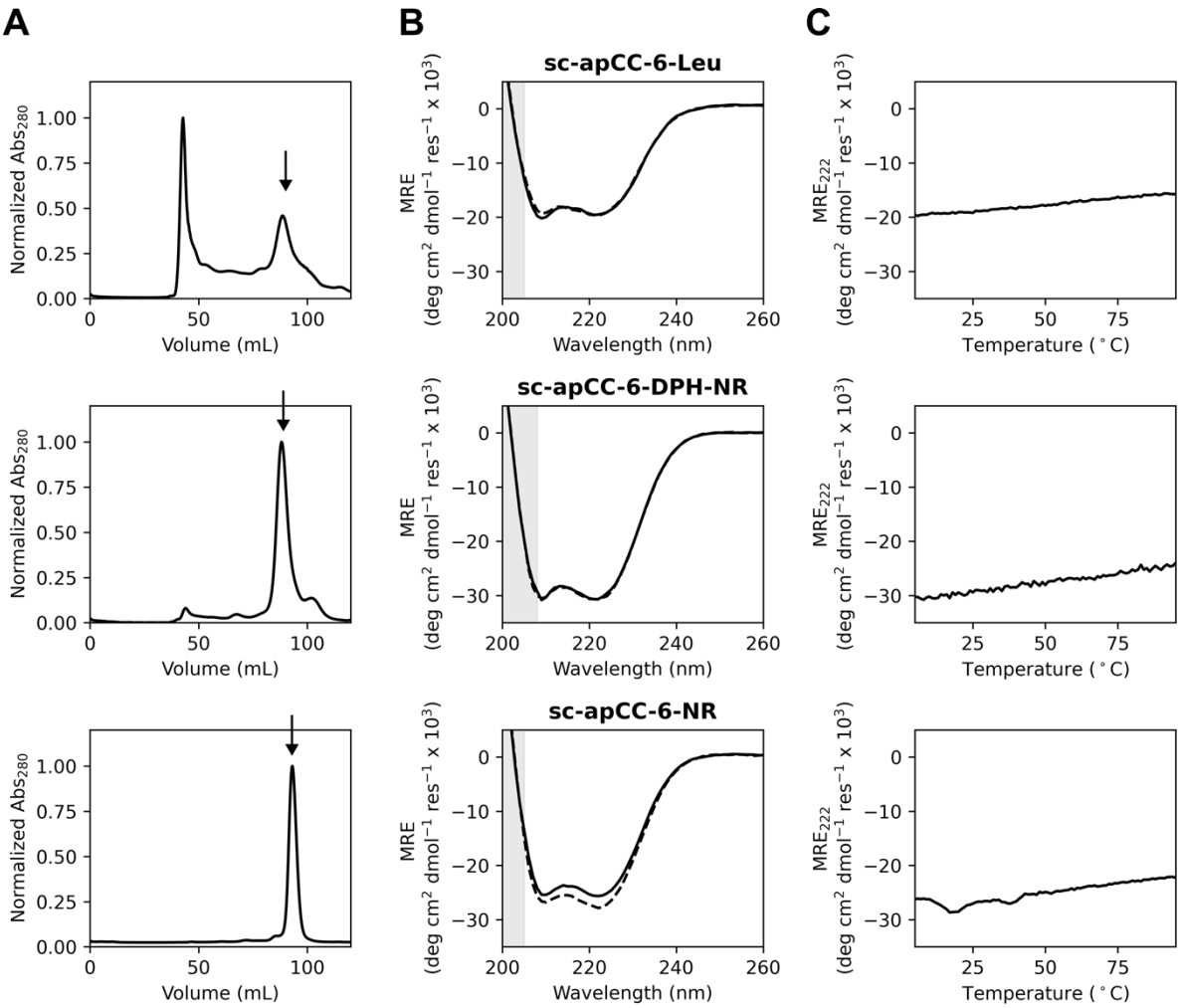
144



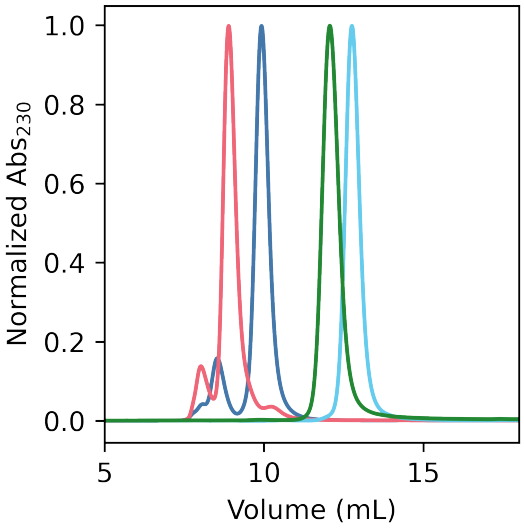
Supplementary Figure 12. Biophysical characterization of apo sc-apCC-4 variants. **(A)** Preparative size-exclusion chromatography (SEC) elution profiles. Proteins were purified on either Superdex 75 pg or Superdex 200 pg columns in 50 mM sodium phosphate, pH 7.4, with 150 mM NaCl. The arrow marks the elution fractions selected for further analysis. **(B)** Far-UV circular dichroism (CD) spectral scans collected at 5 °C (solid lines) and after heating to 95 °C followed by cooling to 5 °C (dashed lines). Filled grey regions indicate HT voltage ≥600 V. **Conditions**: 10 µM protein in 50 mM sodium phosphate, 150 mM NaCl, pH 7.4. **(C)** Temperature-dependent CD monitored at 222 nm. Thermal melts were recorded on heating (solid lines) and on cooling (dashed lines) to/from 95 °C. **Conditions**: 10 µM protein in 50 mM sodium phosphate, 150 mM NaCl, pH 7.4.



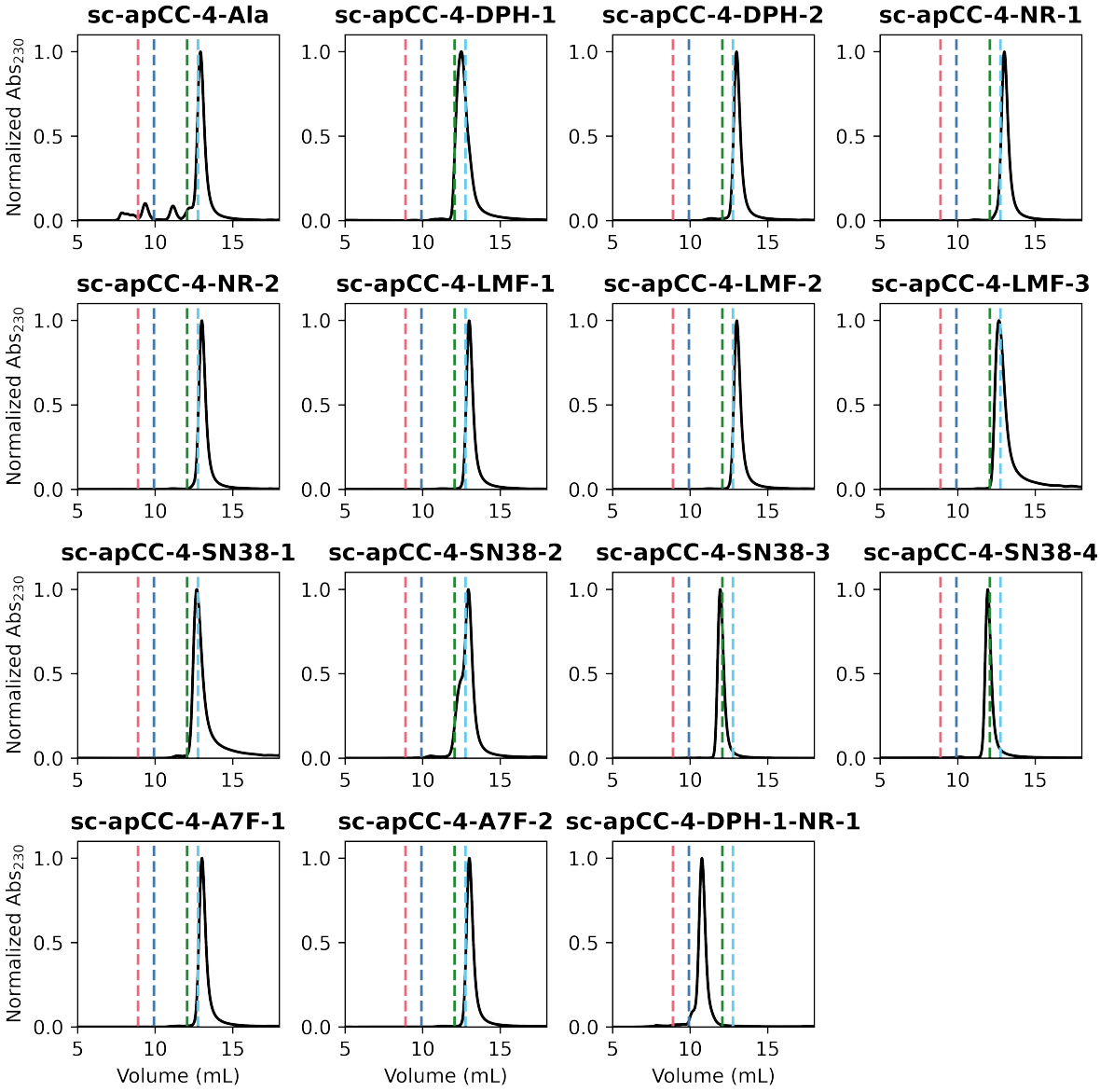
Supplementary Figure 13. Biophysical characterisation of sc-apCC-4-SN38-1 at pH 4. (A) Far-UV circular dichroism (CD) spectral scans collected at 5 °C (solid lines) and after heating to 95 °C followed by cooling to 5 °C (dashed lines). Filled grey regions indicate HT voltage ≥600 V. **(B)** Temperature-dependent CD monitored at 222 nm. Thermal melts were recorded on heating (solid lines) and on cooling (dashed lines) to/from 95 °C. **Conditions**: 10 μM protein in 100 mM sodium acetate, 150 mM NaCl, pH 4.



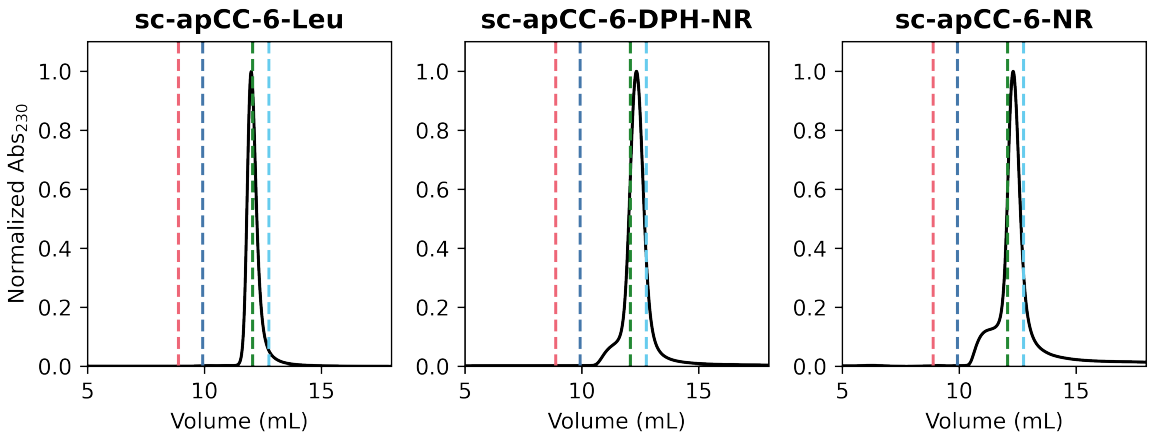
Supplementary Figure 14. Biophysical characterization of apo sc-apCC-6 variants. **(A)** Preparative size-exclusion chromatography (SEC) elution profiles. Proteins were purified on either Superdex 75 pg or Superdex 200 pg columns in 50 mM sodium phosphate, pH 7.4, with 150 mM NaCl. The arrow marks the elution fractions selected for further analysis. **(B)** Far-UV circular dichroism (CD) spectral scans collected at 5 °C (solid lines) and after heating to 95 °C followed by cooling to 5 °C (dashed lines). Filled grey regions indicate HT voltage ≥600 V. **Conditions**: 10 µM protein in 50 mM sodium phosphate, 150 mM NaCl, pH 7.4. **(C)** Temperature-dependent CD monitored at 222 nm. Thermal melts were recorded on heating (solid lines) and on cooling (dashed lines) to/from 95 °C. **Conditions**: 10 µM protein in 50 mM sodium phosphate, 150 mM NaCl, pH 7.4.



Supplementary Figure 15. Calibration curve used for aFPLC analysis. Conalbumin (76 kDa), red. Ovalbumin (44 kDa), blue. Sc-apCC-6-SLLA (26 kDa, PDB id: 8QAE), green. Sc-apCC-4 (17 kDa, PDB id: 8A3K), cyan. **Conditions:** 50 mM sodium phosphate, pH 7.4, 150 mM NaCl. Proteins run on Superdex 75 Increase 10/300 GL at 0.5 – 2 mg/ml.



Supplementary Figure 16. aFPLC traces of all 4-helix bundle proteins in this study. Dashed lines represent elution volumes of calibration standards (Figure S13). **Conditions:** 50 mM sodium phosphate, pH 7.4, 150 mM NaCl. Proteins run on Superdex 75 Increase 10/300 GL at 0.5 – 2 mg/ml.



185

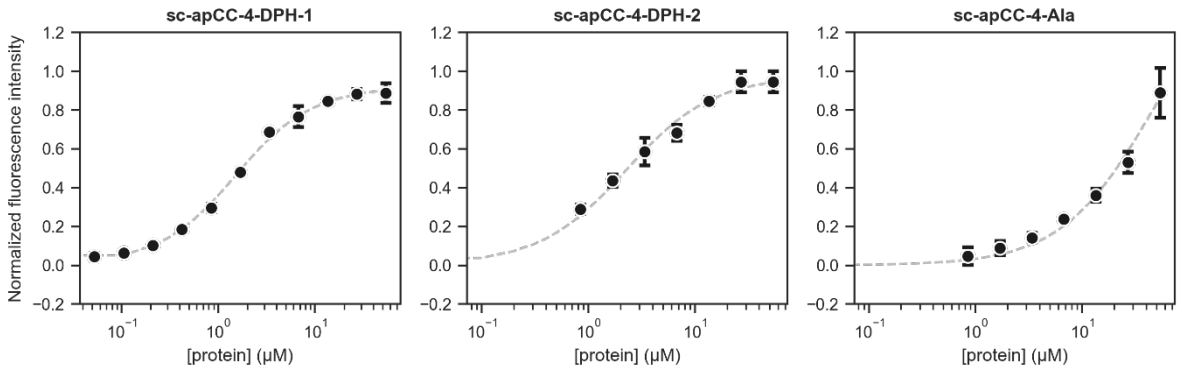
186 **Supplementary Figure 17. aFPLC traces of all 6-helix bundle proteins in this study.** Dashed lines
187 represent elution volumes of calibration standards (Figure S13). **Conditions:** 50 mM sodium
188 phosphate, pH 7.4, 150 mM NaCl. Proteins run on Superdex 75 Increase 10/300 GL at 0.5 – 2 mg/ml.

189

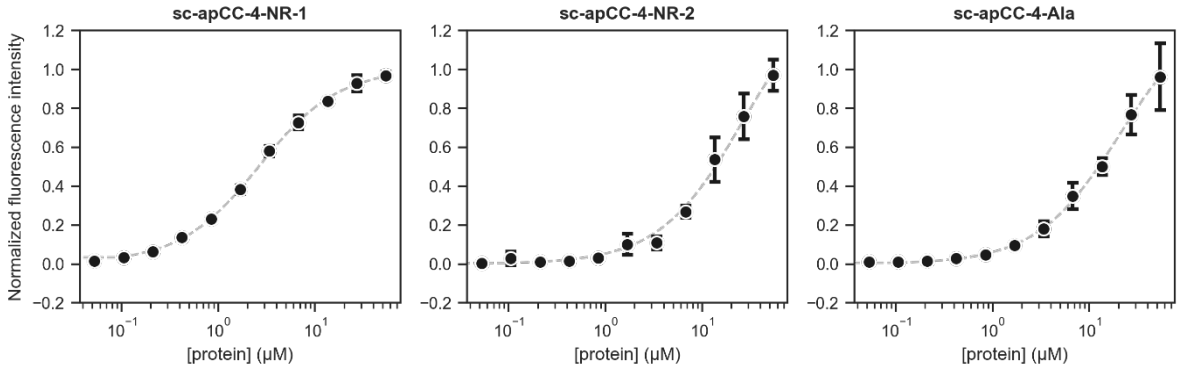
190

191

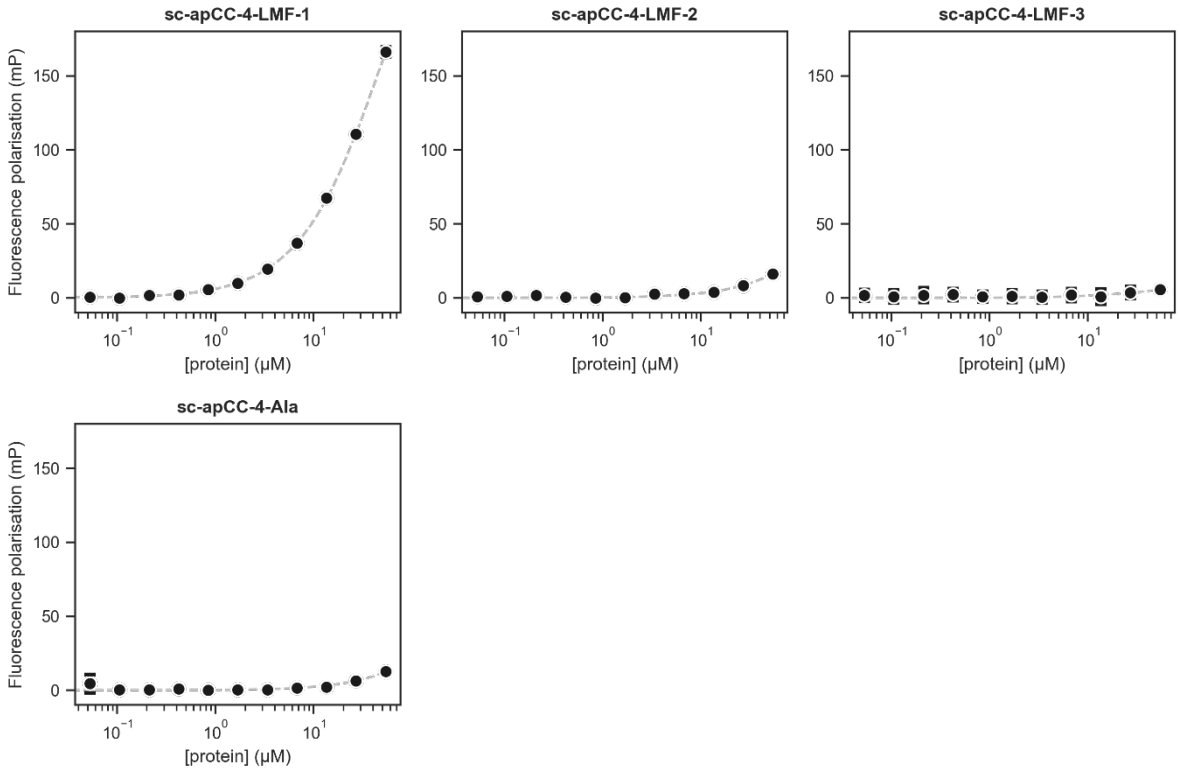
2.2 Characterisation of 4-helix binders



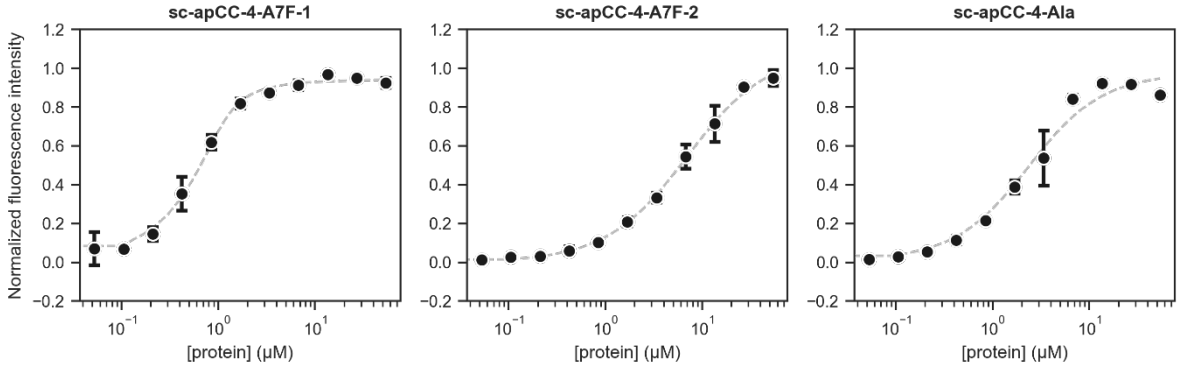
Supplementary Figure 18. Saturation binding curves for 4-helix bundle DPH binders. Fits returned $K_{D(sc\text{-}apCC\text{-}4\text{-}DPH\text{-}1)}$ = 1.3 ± 0.1 μM, $K_{D(sc\text{-}apCC4\text{-}DPH\text{-}2)}$ = 2.0 ± 0.2 μM, $K_{D(sc\text{-}apCC\text{-}4\text{-}Ala)}$ > 50 μM**.** **Conditions**: 0.5 μM ligand, 0 – 60 μM protein, data are the mean of three independent repeats, error bars represent the standard deviation from the mean. All data collected in PBS, 10% v/v MeCN, pH 7.4.



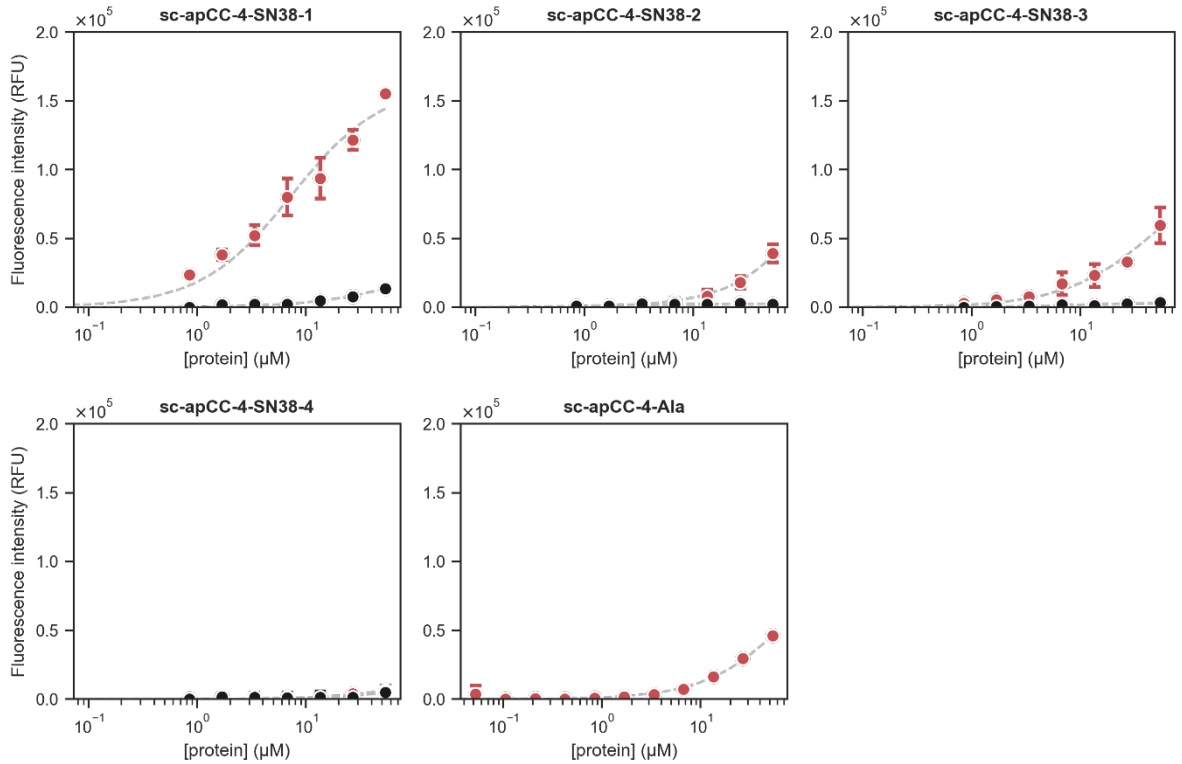
Supplementary Figure 19. Saturation binding curves for 4-helix bundle Nile red binders. Fits returned $K_{D(sc\text{-}apCC\text{-}4\text{-}NR\text{-}1)}$ = 2.4 ± 0.1 μM, $K_{D(sc\text{-}apCC\text{-}4\text{-}NR\text{-}2)}$ > 25 μM, $K_{D(sc\text{-}apCC\text{-}4\text{-}Ala)}$ > 20 μM**.** **Conditions**: 0.5 μM ligand, 0 – 60 μM protein, data are the mean of three independent repeats, error bars represent the standard deviation from the mean. All data collected in PBS, 10% v/v MeCN, pH 7.4.



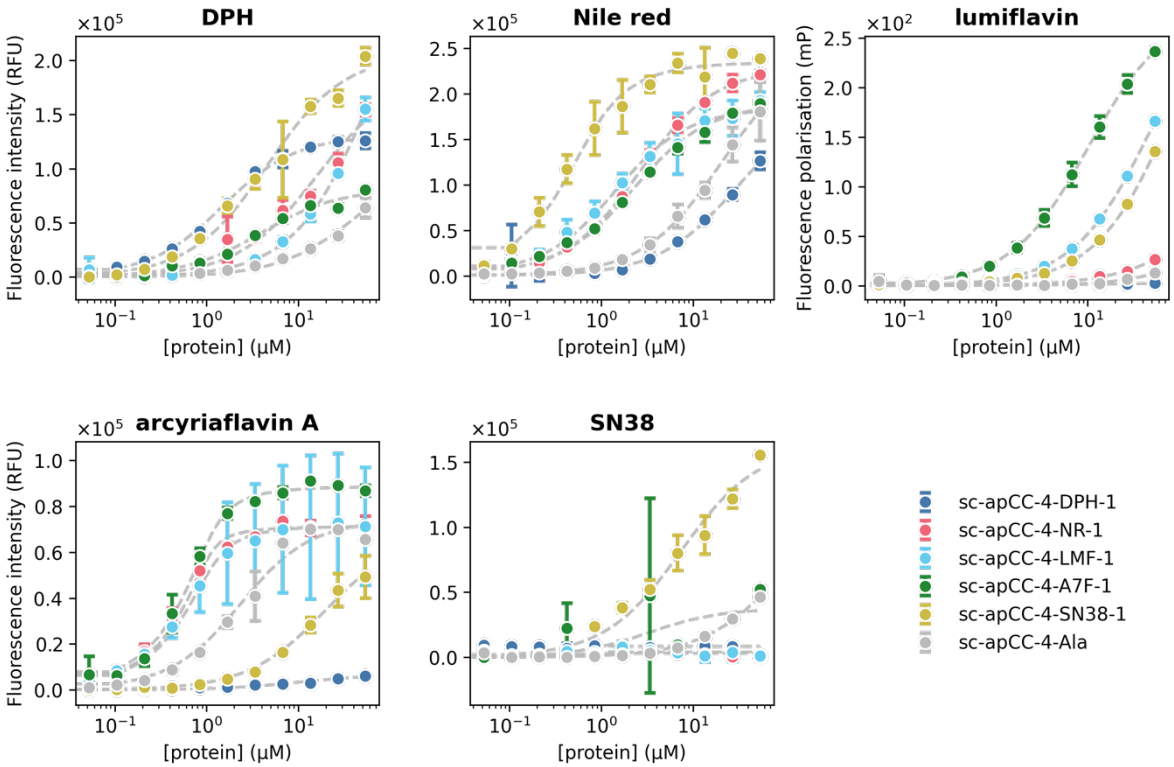
Supplementary Figure 20. Saturation binding curves for 4-helix bundle lumiflavin binders. Fit returned $K_{D(sc\text{-}apCC\text{-}4\text{-}LMF\text{-}1)}$ > 50 μM, sc-apCC-4-LMF-2, sca-apCC-4-LMF-3 and sc-apCC-4-Ala showed no binding. **Conditions**: 1 μM ligand, 0 – 60 μM protein, data are the mean of three independent repeats, error bars represent the standard deviation from the mean. All data collected in PBS, 10% v/v DMSO, pH 7.4.



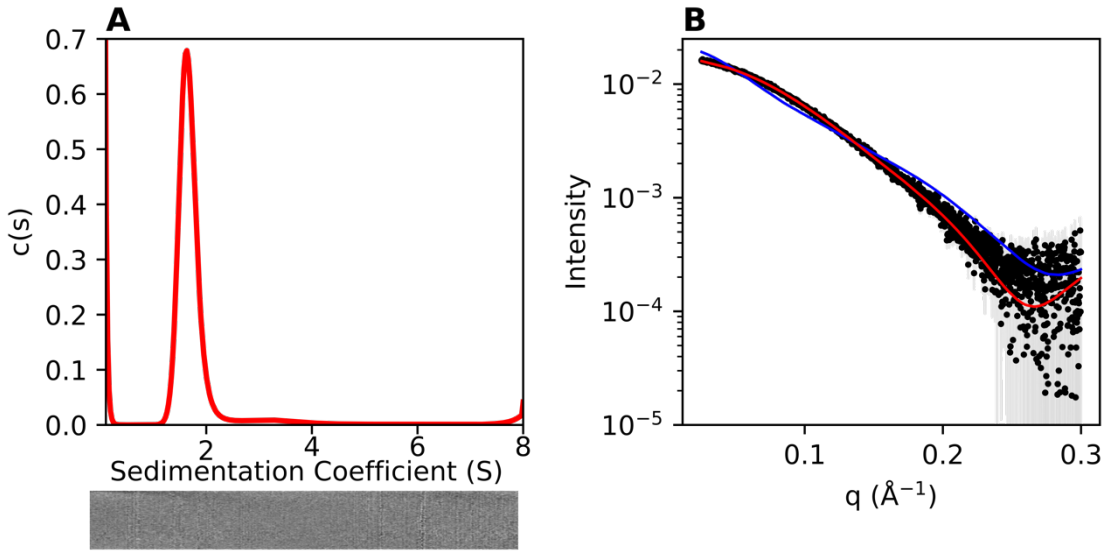
Supplementary Figure 21. Saturation binding curves for 4-helix bundle arcyriaflavin A binders. Fits returned $K_{D(sc\text{-}apCC\text{-}4\text{-}A7F\text{-}1)}$ = 0.11 ± 0.02 μM, $K_{D(sc\text{-}apCC\text{-}4\text{-}A7F\text{-}2)}$ = 6.5 ± 0.3 μM, $K_{D(sc\text{-}apCC\text{-}4\text{-}Ala)}$ = 1.8 ± 0.3 μM. **Conditions**: 1 μM ligand, 0 – 60 μM protein, data are the mean of three independent repeats, error bars represent the standard deviation from the mean. All data collected in PBS, 10% v/v MeCN, pH 7.4.



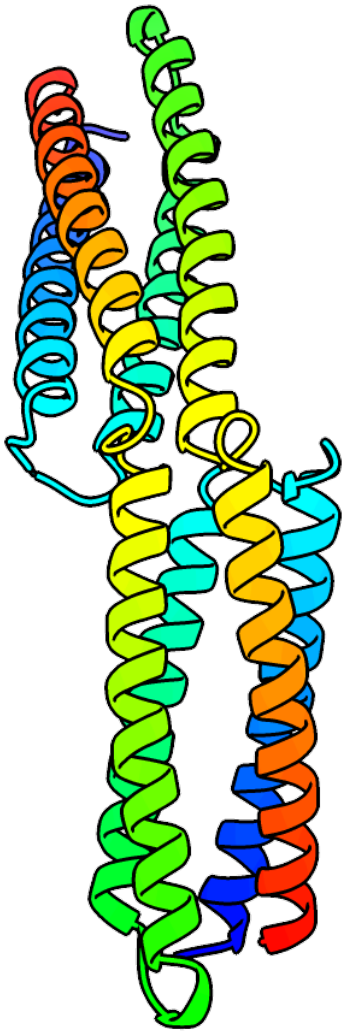
Supplementary Figure 22. Saturation binding curves for 4-helix bundle SN38 binders measured at pH 7.4 (black) or pH 4 (red). At pH ≤ 4.5, SN38 lactone ring is stable (our design target), resulting in stronger binding, whereas at pH 7.4 the ring undergoes reversible opening. Fits for pH 4 returned $K_{D(sc\text{-}apCC\text{-}4\text{-}SN38\text{-}1)}$ > 5 μM. **Conditions**: 1 μM ligand, 0 – 60 μM protein, data are the mean of three independent repeats, error bars represent the standard deviation from the mean. All data collected either in ABS, 5% v/v DMSO, pH 4 (red) or PBS, 5% v/v DMSO, pH 7.4 (black).



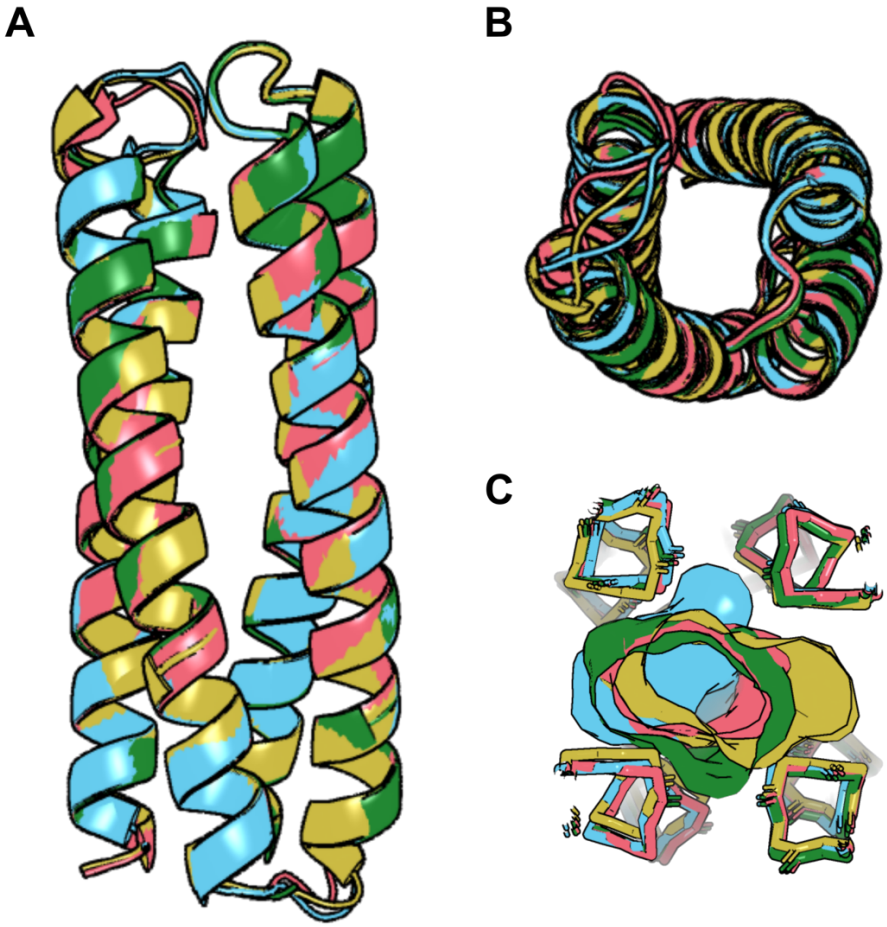
Supplementary Figure 23. Saturation binding curves showing ligand selectivity for the highest-affinity designs in sc-apCC-4. Conditions: 0.5 μM DPH/Nile red or 1 μM lumiflavin/arcyriaflavinA/SN38, 0 – 60 μM protein, data are the mean of three independent repeats, error bars represent the standard deviation from the mean. All data were collected in PBS, pH 7.4 with either 10% v/v MeCN (DPH/Nile red/aryriaflavin A) or 10% v/v DMSO (lumiflavin/SN38). See Supplementary Table 5 for fitted binding affinities.



Supplementary Figure 24. A) Sedimentation velocity and (B) small-angle X-ray scattering analysis of the sc-apCC-4-DPH-2. Sedimentation velocity fit returned MW = 17.6 kDa (1x monomer), SAXS fits returned χ^2 **= 1.7** for a monomer (red) and χ^2 **= 33.8** for the dimer (blue). **Sedimentation velocity conditions**: 40 000 RPM, 30 µM protein, 50 mM sodium phosphate and 150 mM NaCl (pH 7.4**). SAXS conditions**: 20 mg/ml protein in 20 mM Tris pH 8.0, 50 mM NaCl, 1% v/v glycerol.

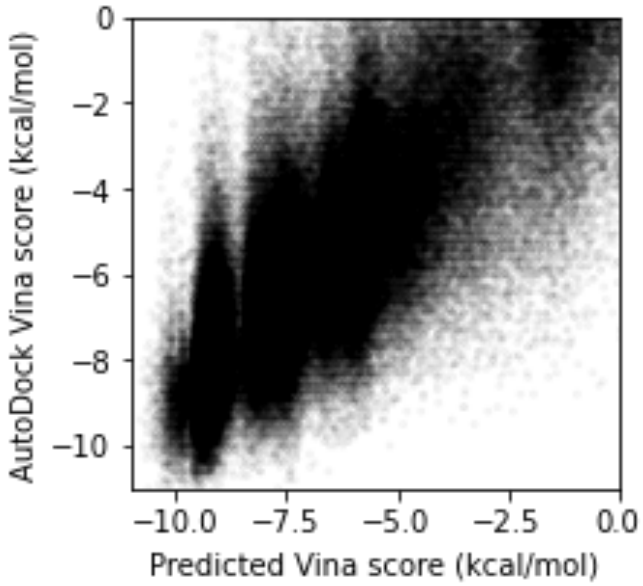


Supplementary Figure 25. The helix-swapped DPH binder, sc-apCC-4-DPH-2 at 2.4 Å resolution (PDB id 9R1N). Two chains are coloured as chainbow, transitioning from blue at the N-terminus to red at the C-terminus.

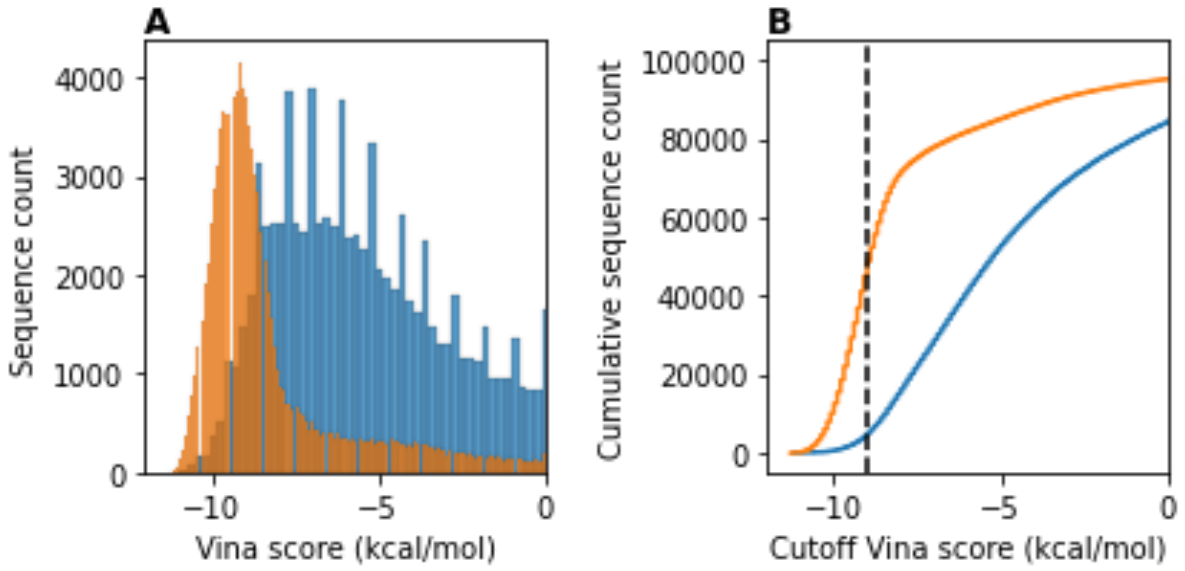


Supplementary Figure 26. Superimposition of the backbones of the best binders (PDB IDs 9R1J, 9R1M, 9R1K, and 9R1N). Cartoon representations of the superimposed backbones shown from the side **(A)** and from the top **(B). (C)** Backbone shown as sticks, with a slice through the centre of the binding cavity. Structures are coloured as in Figure 2. The DPH binder is omitted due to its domain-swapped architecture.

2.3 Design and characterisation of 6-helix binders

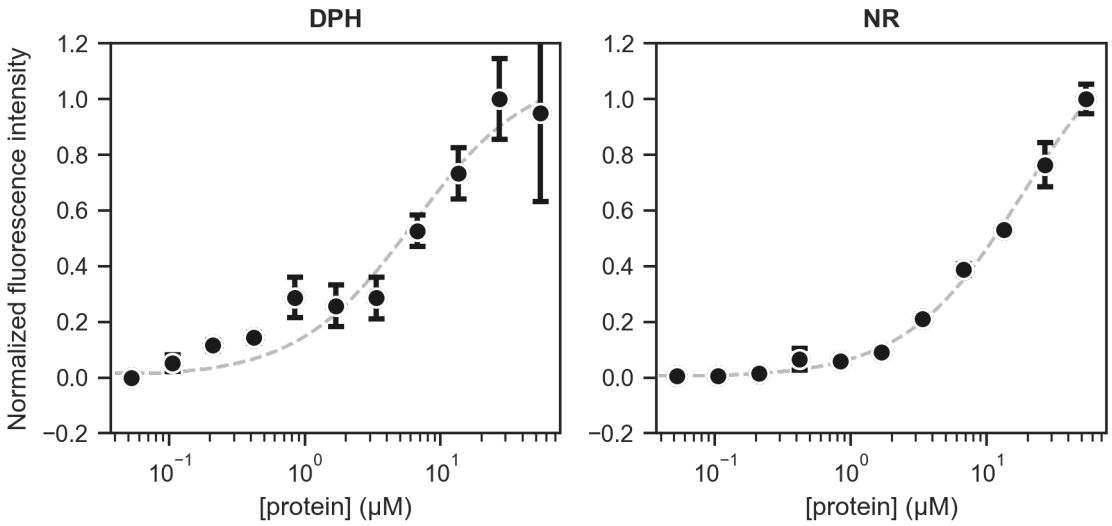


Supplementary Figure 27. Random Forest predictor performance on a test set (100 000 sequences) after actively sampling 50 000 sequences from a full set of 1 277 192 sequences. RMSE = 2.29 kcal/mol.

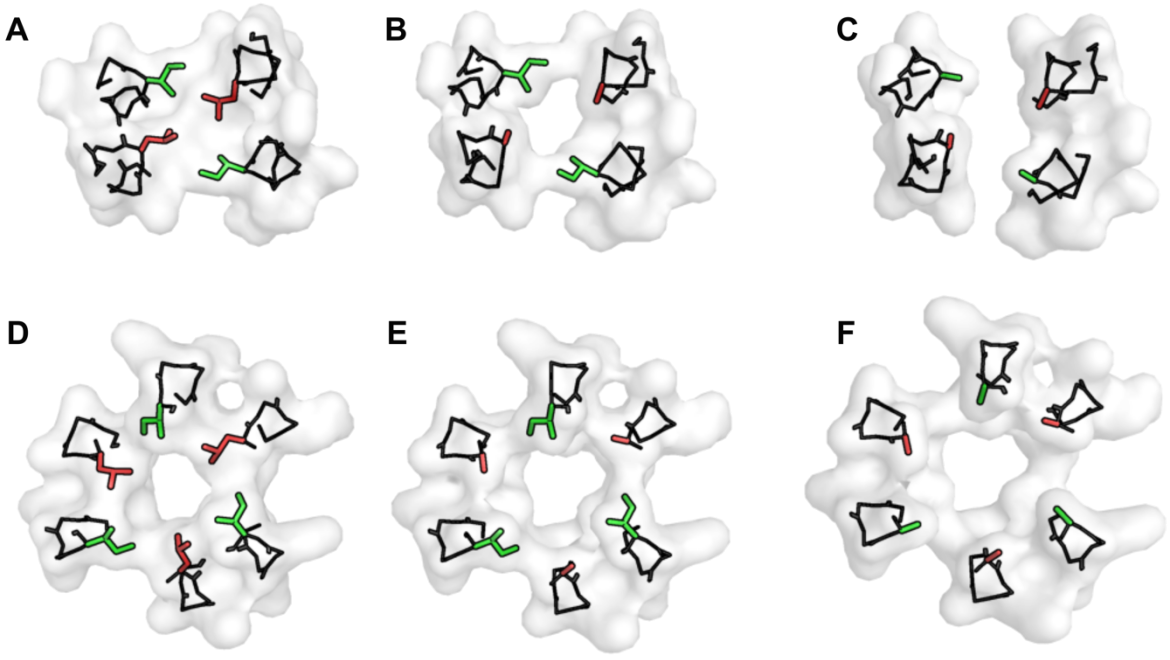


Supplementary Figure 28. Comparison between random sampling (blue) versus active sampling (orange). (A) AutoDock Vina score distribution in each dataset, 100 000 sequences each from a full set of 1 277 192 sequences. (B) Cumulative distribution plot showing the number of sequences with docking scores less or equal to the cutoff. At -9 kcal/mol cutoff, we observe ~11 fold increase in

sequence number with active sampling (44 988 sequences) compared to random sampling (4189 sequences).



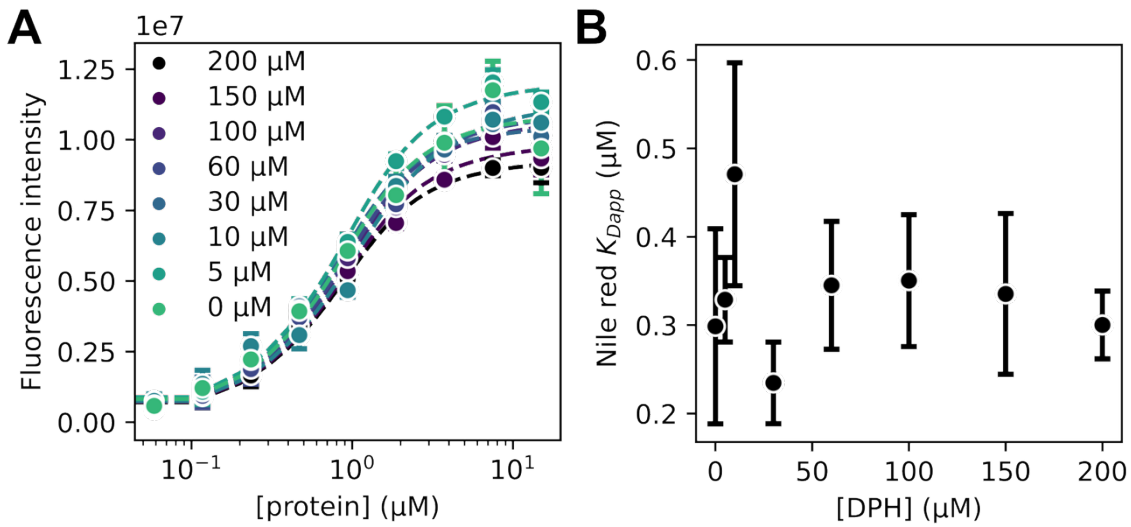
Supplementary Figure 29. Saturation binding curves for the parent 6-helix bundle (all Leu). Nile red, orange; DPH, teal. $K_{D(DPH)}$ = 6.1 ± 1.9 μM, $K_{D(NR)}$ > 20 μM. **Conditions**: 0.5 μM, 0 – 60 μM protein, data are the mean of three independent repeats, error bars represent the standard deviation from the mean. All data collected in PBS, 10% v/v MeCN, pH 7.4.



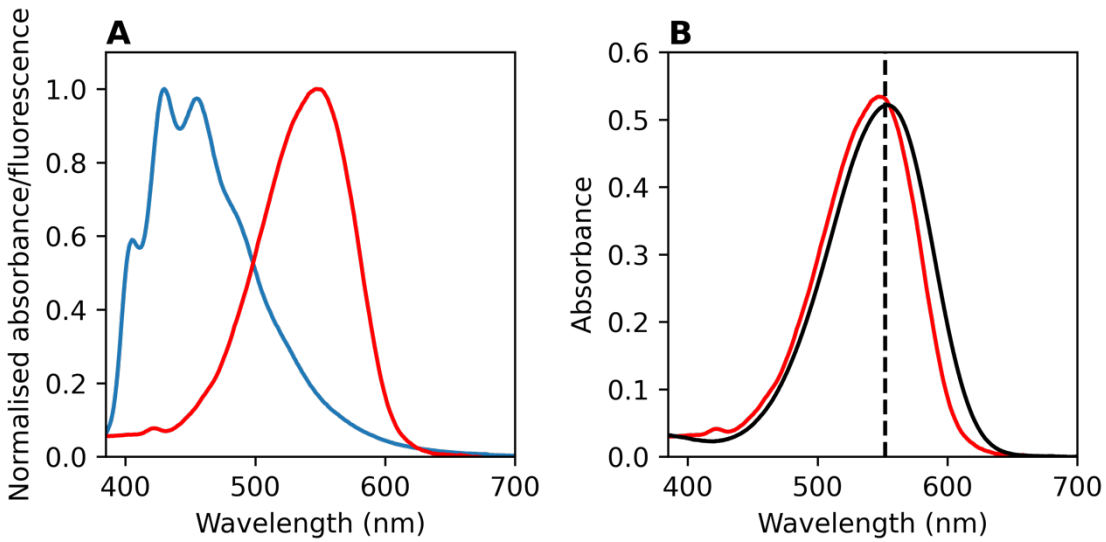
Supplementary Figure 30. Comparison of the effect of modifying *a* (red) and *d* (green) positions in sc-apCC-4 (ABC; PDB id, 8a3k) and sc-apCC-6 (DEF; PDB id, 8qad). A&D, original protein; **B&E**, Leu at ***a*** mutated into Ala; **C&F,** Leu at ***a*** and Ile at ***d*** mutated into Ala. In both cases, mutations at **d**

create side-channels which are necessary to accommodate the ligands in the smaller sc-apCC-4 but not the sc-apCC-6.

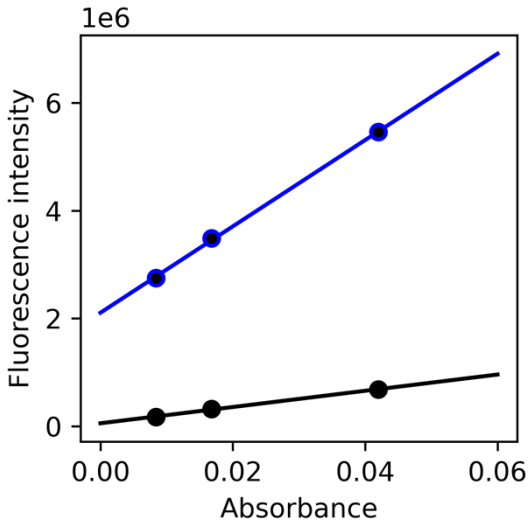
2.4 Spectroscopic characterisation and simulations of the double-binder



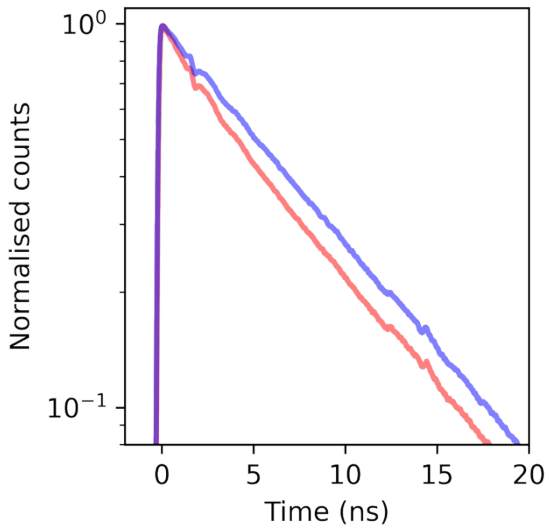
Supplementary Figure 31. Apparent binding affinity of Nile red in the presence of varying DPH concentrations. **(A)** Titration curves showing Nile red fluorescence intensity. **(B)** Fitted apparent K_D values for Nile red in presence of DPH. **Conditions**: 1 µM Nile red, 0 – 15 µM protein, 0 – 200 µM DPH, data are the mean of two independent repeats, error bars represent the standard deviation from the mean. All data collected in PBS, 10% v/v MeCN, pH 7.4.



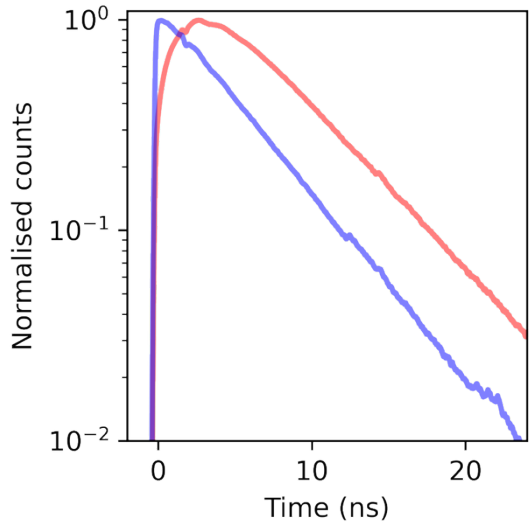
Supplementary Figure 32. Spectral overlap integral (7.79 x 10^{14} nm^4 M^{-1} cm^{-1}) calculation between DPH and Nile red in the double binder. (A) Normalised DPH emission (blue) and Nile red absorption (red). **(B)** Absorption spectrum for Nile red in the double binder (red) and methanol (black). Molar extinction coefficient for Nile red in methanol (45 000 $M^{-1}cm^{-1}$ at 552 nm; dashed black) was used for the spectral overlap calculation. **Conditions:** 10 µM of either DPH or Nile red, 30 µM protein, PBS, 10% v/v MeCN, pH 7.4 or 10 µM Nile red in methanol.



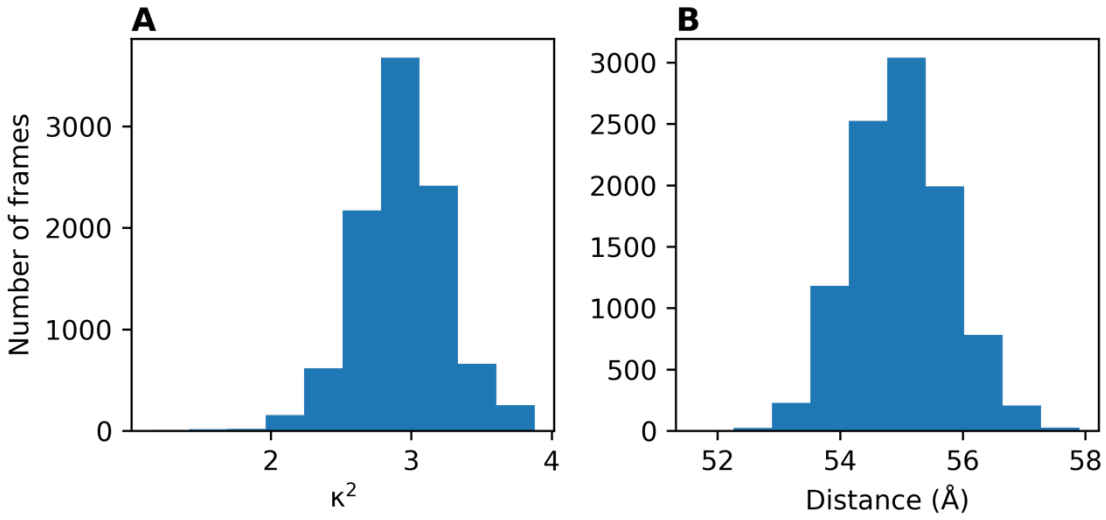
Supplementary Figure 33. Fluorescence quantum yield estimation for DPH in the double binder (blue, Φ = 85%) by comparison to a standard – DPH in MeCN (black, Φ = 17%). Conditions: 30 µM protein, data are the mean of three independent repeats, error bars (standard deviation from the mean) fall within the size of the markers.



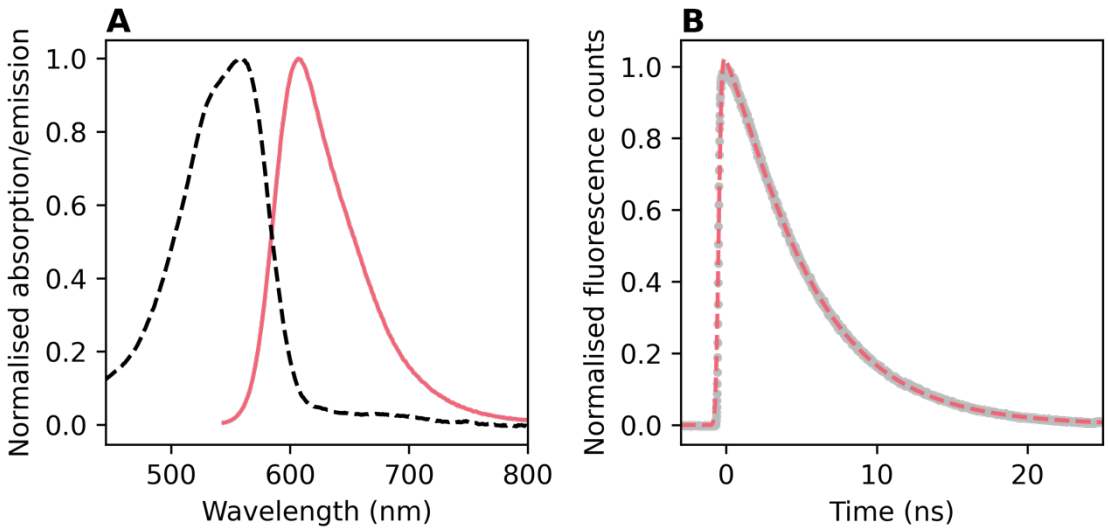
Supplementary Figure 34. Normalised TCSPC data for DPH with (red) and without (blue) Nile red after excitation at 352 nm. For the binder containing only DPH, a mono-exponential fit returned τ(DPH) = 7.6 ± 0.2 ns. **Conditions:** 20 μM of each ligand, 30 μM protein, PBS, 10% v/v MeCN, pH 7.4.



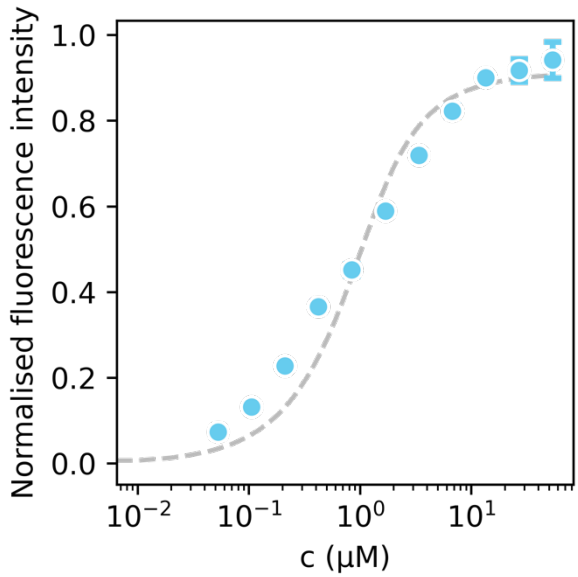
Supplementary Figure 35. Normalised TCSPC data for Nile red with (red) and without (blue) DPH after excitation at 352 nm. Emission from Nile red after direct excitation without DPH was accumulated for 20-times longer to account for the low Nile red absorption at this excitation wavelength. Data fit to a mono-exponential decay returned τ(Nile red) = 5.2 ± 0.2 ns. **Conditions:** 20 μM of each ligand, 30 μM protein, PBS, 10% v/v MeCN, pH 7.4.



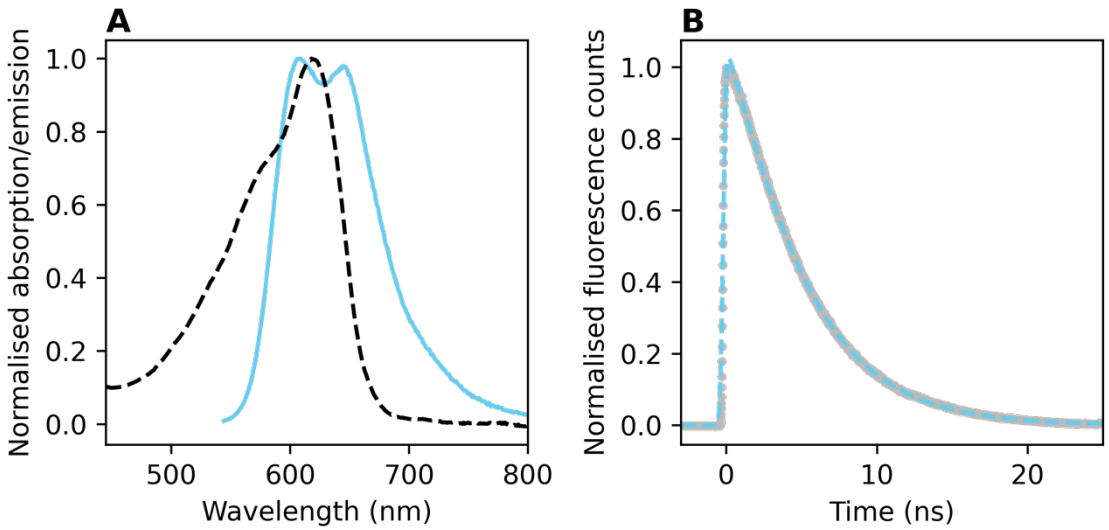
Supplementary Figure 36. (A) Dipole orientation factor (κ^2) and (B) centre of mass DPH – Nile red distances derived from MD simulations. The starting model was generated with AlphaFold2, ligands were docked with AutoDock Vina. 5 independent 100 ns simulations were performed. For both ligands, transition dipoles were taken as the long axes of the molecule. $\kappa^2 = 2.9 \pm 0.3$ was used for ligand distance calculation from the experimental data using the Förster equation.



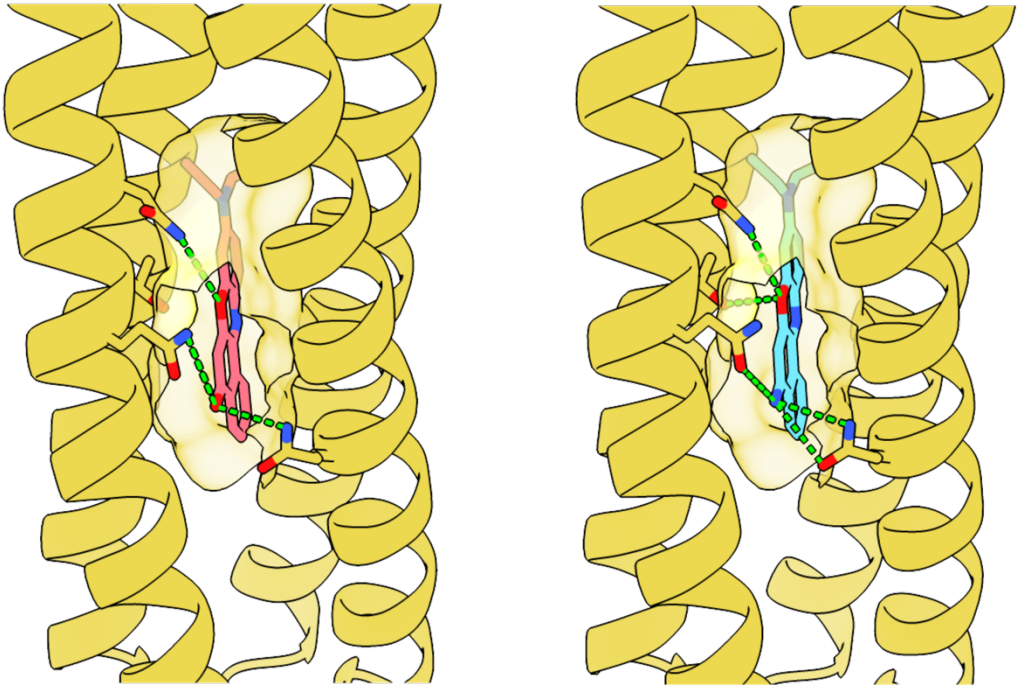
Supplementary Figure 37. Spectral properties of Nile red bound to sc-apCC-4-SN38-1. (A) Absorption (dashed) and emission (solid, λ_{ex} = 510 nm) spectra. (B) Time-correlated single photon counting data showing fluorescence decay after excitation at 560 nm. The data was fitted to a mono-exponential decay with a rise component[2]. τ_{NR} = 4.8 ± 0.2 ns. **Conditions:** 5 µM Nile red, 50 µM protein, PBS, 10% v/v MeCN, pH 7.4.



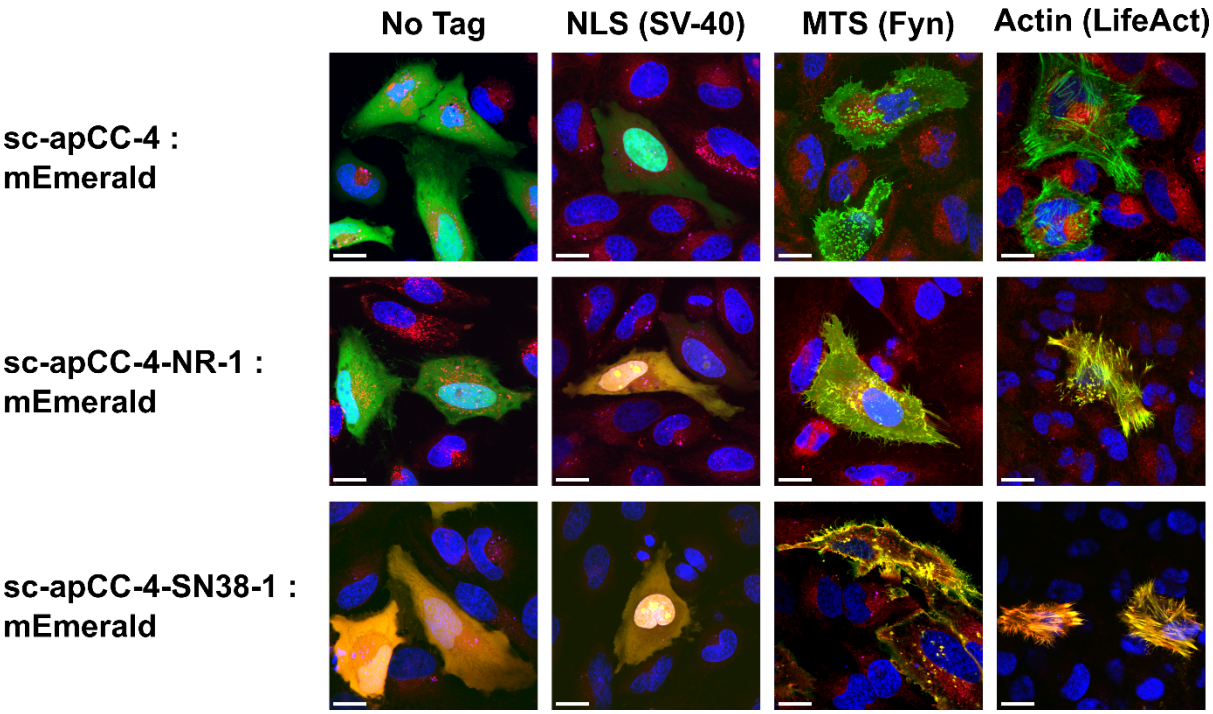
Supplementary Figure 38. Saturation binding data and fits for Nile blue binding to sc-apCC-4-SN38-1. Kd = 0.40 ± 0.13 μM. **Conditions**: 1 μM Nile blue, 0 – 60 μM protein, data are the mean of three independent repeats, error bars represent the standard deviation from the mean.



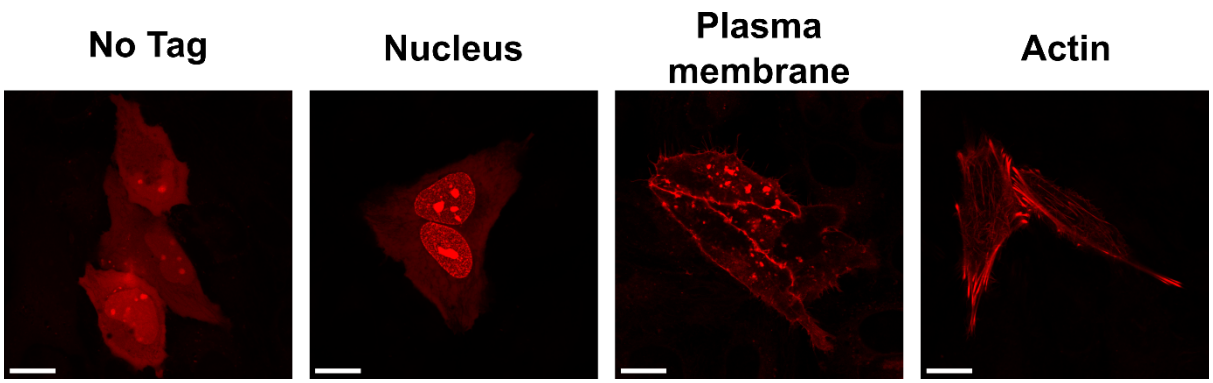
Supplementary Figure 39. Spectral properties of Nile blue bound to sc-apCC-4-SN38-1. (A) Absorption (dashed) and emission (solid, λ_{ex} = 530 nm) spectra. (B) Time-correlated single photon counting data showing fluorescence decay after excitation at 560 nm. The data was fitted to a mono-exponential decay with a rise component[2]. τ_{NR} = 4.2 ± 0.2 ns. **Conditions:** 5 μM Nile blue, 50 μM protein, PBS, 10% v/v MeCN, pH 7.4.



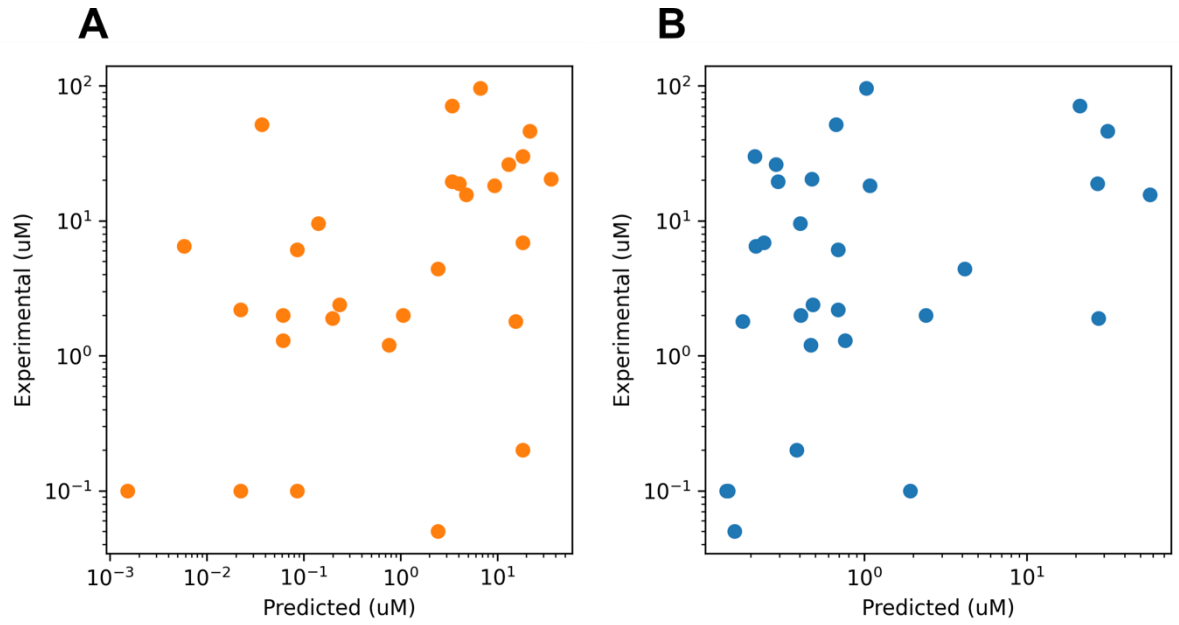
Supplementary Figure 40. Nile red (red) and Nile blue (cyan) docked into the crystal structure of sc-apCC-4-SN38-1 (PDB id 9R1L). Nile red Vina score: -10.4 kcal/mol, Nile blue Vina score: -10.6 kcal/mol. Sc-apCC-4-SN38-1 structure was energy-minimised using Rosetta prior to docking.



Supplementary Figure 41. Fluorescence confocal microphotographs of live HeLa cells expressing the non-binding sc-apCC-4 scaffold (top row) or the designer binders sc-apCC-4-NR-1 (middle row), sc-apCC-4-SN38-1 (bottom row) as fusions to mEmerald and stained with Nile red and Hoechst 33342. Cytosolic constructs as well as fusions to short sequences targeting the nucleus, plasma membrane, and actin are presented in thedifferent columns. The composite images show mEmerald (green: ex/em 488/500-550 nm), NR (red: ex/em 561/580-654 nm), and Hoechst 33342 (blue: ex/em 405/417-477 nm) channels. Scalebar = 20 µm.



Supplementary Figure 42. Fluorescence confocal microphotographs of live HeLa cells expressing sc-apCC-4-SN38-1 fused to short N-terminal sequences targeting the nucleus, plasma membrane, and actin, but without the mEmerald domain. Images were obtained 30 minutes after staining with 0.1 µM Nile red added to the cell media and then observing Nile red in the red channel (ex/em 561/580-654 nm). Scalebar = 20 µm.



Supplementary Figure 43. Correlation of experimental and predicted binding constants. **(A)** Affinities calculated from AutoDock Vina docking scores using the AlphaFold2 models for all experimentally validated ligand–binder pairs. **(B)** Predicted affinities calculated from Boltz-2 structure predictions with the ligand. Since both in silico methods produce binding energies, predicted affinities were derived using the relationship $\Delta G = RT \ln K_D$.

Supporting references

1 Menzer, W. M., Xie, B. & Minh, D. D. L. On Restraints in End-Point Protein–Ligand Binding Free Energy Calculations. *J. Comput. Chem.* **41**, 573-586 (2020).

2 Gajo, C. *et al.* Nile Red Fluorescence: Where's the Twist? *J. Phys. Chem. B* **128**, 11768-11775 (2024).